The Dissertation Committee for Drew Taylor Wagner Certifies that this is the approved version of the following dissertation:

Structural and Functional Studies of Uncommon Polyketide Synthase Domains

Committee:

Adrian Keatinge-Clay, Supervisor

Hal Alper

David Hoffman

Hung-Wen Liu

Edward Marcotte

Structural and Functional Studies of Uncommon Polyketide Synthase Domains

by

Drew Taylor Wagner

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2017

This is for my mom and dad.

Acknowledgements

Well, I can't believe I'm here already. I suppose a round of "thank yous" are in order for the mere fact that I got here at all but also because I actually enjoyed most of this whole grad school thing. I should start by thanking the various Keatinge-Clay lab members and collaborators I have had the distinct pleasure of working with over the past several years because without them this work would have been impossible. My lab mates in particular have been such a rich source of knowledge, insight and joy, enriching not only my grad school experience but also me as a person. In no particular order, thank you Dr. Cole Stevens for being an all-around solid human being, Dr. Chris Fage for your patience and support, Dr. Darren Gay for teaching me experimental crystallography and what a proper choke-hold feels like, Dr. Connie Bailey for your encyclopedic knowledge of polyketides, Dr. Shawn Piasecki for the baller tan, Hannah Manion for purifying like 300 proteins, Alex Zhang for not setting the lab on fire, Jessica Meinke for bossing me around and always picking the worst papers for group meeting and Dr. Amanda Jane Hughes-Brenes for just being hilarious and also for that song guessing game we used to play. Also thank you to my advisor, Adrian Keatinge-Clay for sharing your love of polyketides with me, you have converted me, I too now love polyketides, though admittedly not as much as you. Much love to you all, lab would not have been the same without you. Also thank you to Rachel Mehaffey and Dr. Jennifer Brodbelt for a fruitful collaboration. Additional shout outs to Andrew T. Nelson, Nick Brenes, Amanda Vaughn and Groves Dixon for all the fun times. It went by way too quick. And lastly, I would be remiss if I didn't thank my parents for attempting to raise me as a compassionate and

disciplined human, I know I didn't always make it easy but you didn't give up and your efforts were not entirely in vain.

Structural and Functional Studies of Uncommon Polyketide Synthase Domains

Drew Taylor Wagner, PhD

The University of Texas at Austin, 2017

Supervisor: Adrian Keatinge-Clay

Over the past 4 billion years or so, beginning in the primordial soup from which terrestrial life descends, Mother Nature has been molding matter into a seemingly endless number of combinations and searching among them for useful properties. Almost certainly, the vast majority of these atomic combinations have held little value to Her but every so often, a particular combination is found, a gem of perfect dimension or reactivity possessing an extraordinary quality. The quality imparts a value to this particular arrangement of matter and because of it the combination is allowed to persist along with all the other extraordinary combinations.

Humanity has long coveted Mother Nature's process and aspired to emulate Her. Counted among Nature's wondrous creations are the polyketides and the eponymous polyketide synthase enzymes that assemble them. As a class of natural products, polyketides display incredible properties and have found wide-spread use in society as antibiotics, cholesterol medications and cancer therapeutics. The enzymatic molecular machines that biosynthesize these miracles are just as impressive and are regarded with envy by human chemists for their ability to make sophisticated chemical agents with unrivaled precision under the mildest of conditions. This work aims to inch forward our understanding of these exquisite assembly lines so that we may better understand and imitate Nature.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Since the identification of the 6-deoxyerythronolide B (6-dEB) polyketide synthase (PKS) almost 30 years ago, it has served as the core paradigm for understanding the structural and mechanistic principles governing polyketide assembly lines (Cortes et al., 1990; Donadio et al., 1991; Khosla et al., 2007). Responsible for the biosynthesis of the aglycone precursor of the macrolide antibiotic erythromycin, the 6-dEB synthase is comprised of 29 distinct enzymatic domains grouped into 6 functional enzyme sets termed modules that each perform a 2-carbon extension of the polyketide main chain utilizing a methyl-malonyl-CoA extender unit (Figure 1-1). The subsequent elucidation of the architecture and operation of 6-dEB synthase became the foundational knowledge base of polyketide biology regarding the assembly line logic and modular structure inherent to these polyketide synthases (Khosla et al., 2007). Largely from work on the 6-dEB synthase it was extrapolated that PKS modules are minimally comprised of an acyl-transferase (AT), acyl carrier protein (ACP) and ketosynthase (KS) domain used respectively to catalyze the selection, transfer and condensation of CoA-tethered extender units onto a growing polyketide scaffold. In addition to these core domains, it was observed that PKS modules generally contain a variable number of non-requisite processing domains that confer the majority of the remarkable amount of chemical and functional variability seen within finished polyketide products. The processing domains found within the 6-dEB synthase are 6 ketoreductases (KR), 1 dehydratase (DH) and 1 enoyl-reductase (ER) necessary for the sequential reduction, dehydration and saturation

respectively of the polyketide β-keto position after each module's elongation step. In addition to this, a terminal thioesterase domain (TE) is used for simultaneous cyclization and chain cleavage (Figure 1-1). Within the 6-DEB synthase, the consecutive and predictable action of monofunctional domains produces a clear collinear relationship between synthase and product where by knowledge of one allows determination of the other. This collinearity principle is considered a hallmark of PKS biology and in no small part contributed to the excitement surrounding early PKS engineering efforts as researchers envisioned using recombinant synthases for the production of made-to-order polyketides (Khosla et al., 2007).

Over the past decade, it has become increasingly clear that the organization observed in the erythromycin system is frequently not extendable to other more recently discovered synthases. Indeed, many of the crucial supposed tenets of polyketide biosynthesis ostensibly do not hold true for other systems. This is particularly true for the more recently discovered *trans*-AT synthases that do not obey the relative parsimony of module structure and chemistry seen in the more canonical *cis*-AT systems exemplified by the 6-dEB synthase (Figure 1-2). The classic minimal definition of a module often breaks down for many *trans*-AT synthases where modules are comprised of domains lacking functional KS or AT domains (Piel, 2010). Additionally, though the traditional processing domains used during erythromycin biosynthesis have been well studied, many novel processing domains have since been discovered whose function obfuscates the collinear relationship between enzyme and product (Helfrich, 2016). In light of the existence of these systems, the classical model for polyketide biosynthesis for which the

6-dEB synthase is the prototype is but a simplified lens for viewing PKS structure and function. To gain a more complete understanding of the power of polyketide synthases there must be a focused effort on elucidating the unconventional and varied aspects of these amazing enzymes.

The bulk of my graduate work has gone towards investigating structural and functional aspects of polyketide biosynthesis that are uncommon or divergent compared to the enzymes and associated chemistry found within the classical 6-dEB pathway. As alluded to above, the 6-dEB synthase is a relatively simple example of a polyketide synthase and branching out into divergent systems we enable a fuller appreciation of the true scope and potential of polyketide synthases as well as what they can be engineered to do.

In **Chapter 2** and **Chapter 3**, an alternate route of methyl branch incorporation into polyketides is investigated. Within the 6-dEB system, methyl branches from the macrolide core are introduced via the incorporation of methyl-malonyl-CoA extender units. A less commonly employed method for methyl incorporation in polyketides is through the use of embedded SAM dependent methyltransferase (MT) domains. **Chapter 2** characterizes the in vitro activity of 6 MT domains from the *cis*-AT gephyronic acid biosynthetic gene cluster and unequivocally demonstrates that all 6 of the domains exclusively methylate β-ketoacyl-thioester substrates suggesting that within a modular context the point of operation of MT domains is only immediately after KS mediated condensation.

Chapter 3 extends the work of **Chapter 2** by kinetically characterizing *trans*-AT MT domains from the bacillaene, difficidin and macrolactin pathways. The work reveals that *trans*-AT MTs also are specific for β-ketoacyl substrates but display only modest in vitro kinetic properties when excised from the synthase.

Chapter 4 is a detailed examination of the enigmatic split bimodules endogenous to many *trans*-AT synthases. Contrary to canonical *cis*-AT systems like the erythromycin synthase, many *trans*-AT synthases house modules that are split across adjacent polypeptides. Within *cis*-AT systems, modules are typically contained on a single polypeptide while *trans*-AT synthases frequently contain modules split across two polypeptides. These splits conform to a few general arrangements with the two most common being the type A and B split dehydrating modules that each feature a C-terminal KS domain followed by an N-terminal DH domain. **Chapter 4** explores these mysterious split bimodules from a structural, bioinformatic and functional perspective and highlights several conserved features unique to them focusing the aberrant secondary structure of bimodule ketoreductase and dehydratase domains. The close look at these modules has helped clarify their novel features as well as the chemistry they perform and has implications for the future of pathway annotation and natural product validation.

Chapter 5 details the structural characterization of a novel enzyme, the pyran synthase (PS) domain. PS domains are often used by *trans*-AT synthases for the installation of 5 and 6 membered ether rings into polyketide backbones. The high-resolution crystal structure of a pyran synthase domain from the ninth module of the sorangicin synthase provides insight into the mechanism of action of this domain and

highlights the high homology PS domains share with the ubiquitous DH domain. Structure guided mutagenesis of the PS domain implicates a catalytic mechanism dependent upon a histidine and asparagine catalytic dyad that respectively activate and stabilize a polyketide hydroxyl nucleophile for ring formation.

The assembly-line nature of polyketide synthases combined with the potent bioactivity of their polyketide products makes them very appealing engineering targets for the development of new therapeutics. Though considerable progress has been made in this regard, genetic and combinatorial manipulation of polyketide synthases has been slowed by an imprecise understanding of the complexity of these enzymes and the range of chemistry they catalyze. It has become clear that these intricate enzyme systems are more dynamic and selective than previously appreciated and it is only through the close dissection of them that we can come to understand them enough to harness their full catalytic power for the creation of new polyketide drugs.

In **Chapters 2** and **3** I cloned all constructs, purified all proteins and set up enzyme reactions jointly with the other primary author on the projects. In **Chapter 4** I performed all structural work pertaining to PksKR3 and assisted with bioinformatic analysis. In **Chapter 5**, I performed all experimental work with the exception of substrate synthesis. In all chapters I contributed to data analysis and manuscript preparation.

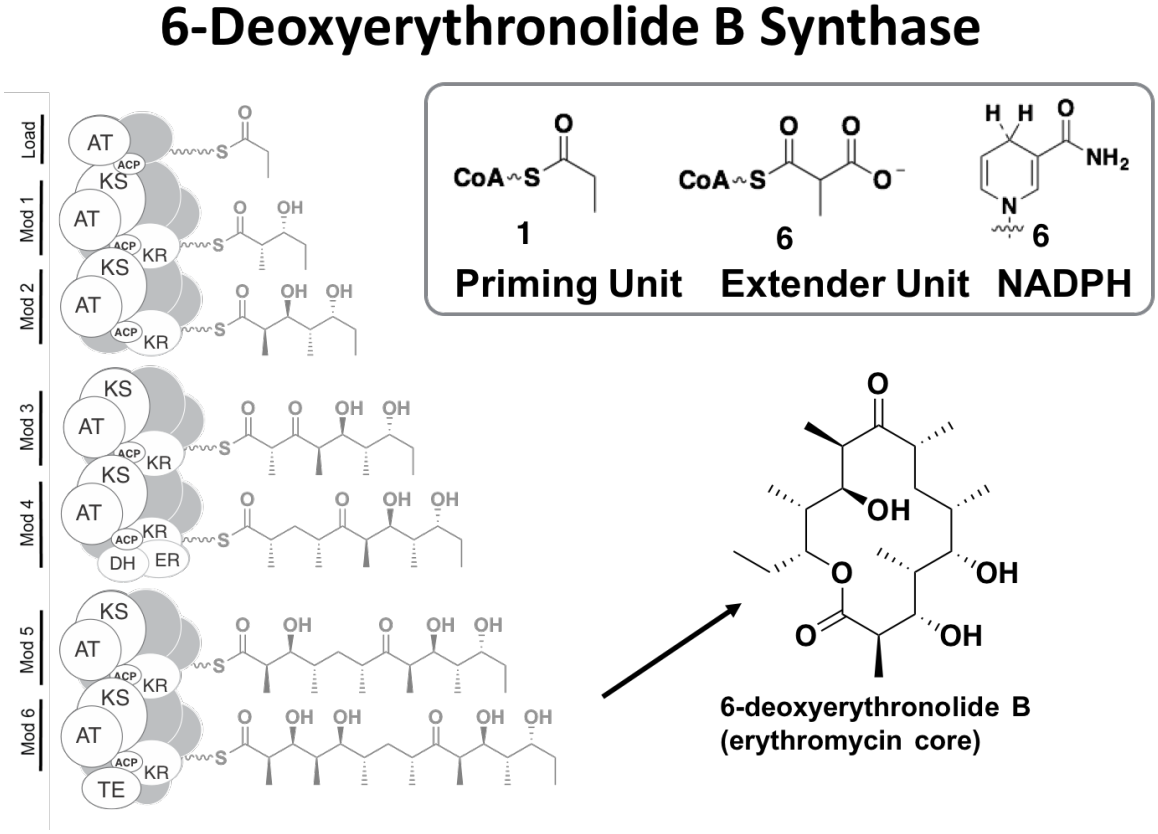


Figure 1-1: The 6-Deoxyerythronolide B Synthase.

6-dEB forms the scaffold of the polyketide antibiotic erythromycin. 6-dEB is produced by a polyketide synthase endogenous to *Saccharopolyspora erythraea*. The synthase consists of six modules that each perform a two carbon extension of the polyketide backbone using a methyl-malonyl-CoA derived extender unit. Each module is defined by a ketosynthase: KS, acyltransferase: AT and acyl carrier protein: ACP that mediate polyketide chain extension. Each module also contains a variable set of processing domains: ketoreductase: KR, dehydratase: DH, enoyl-reductase: ER.

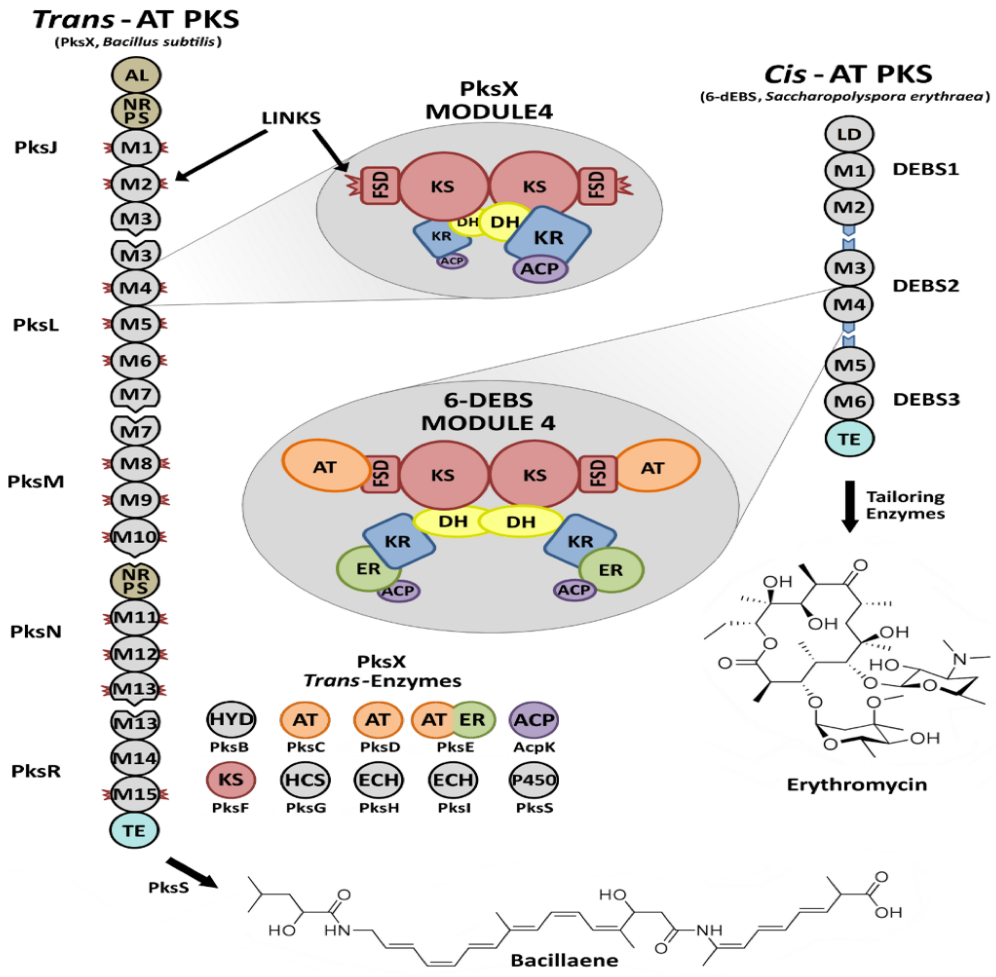


Figure 1-2: *Cis*-AT vs *trans*-AT Polyketide Synthases.

In comparison to canonical *cis*-AT polyketide synthases such as the 6-dEB synthase, *trans*-AT systems are much more varied both in the organization of their modules as well as the chemistry they employ. The rules governing *trans*-AT polyketide synthases are not as well understood as those governing their simpler *cis*-AT counterpart.

Chapter 2: α-Methylation Follows Condensation in the Gephyronic Acid Modular Polyketide Synthase

Adapted from Wagner D, et al. (2016) *Chem Commun (Camb)*. **11**, 2466-74.

ABSTRACT

C-methyltransferases (MTs) from modular polyketide synthase assembly lines are relatively rare and unexplored domains that are responsible for installing α-methyl groups into nascent polyketide backbones. The stage at which these synthase-embedded enzymes operate during polyketide biosynthesis has yet to be conclusively demonstrated. In this work we establish the activity and substrate preference for six MTs from the gephyronic acid polyketide synthase and demonstrate their ability to methylate both *N*-acetylcysteamine- and acyl carrier protein-linked β-ketoacylthioester substrates but not malonyl thioester equivalents. These data strongly indicate that MT-catalyzed methylation occurs immediately downstream of ketosynthase-mediated condensation during polyketide assembly. This work represents the first successful report of MT-catalyzed mono- and dimethylation of simple thioester substrates and provides the groundwork for future mechanistic and engineering studies on this important but poorly understood enzymatic domain.

INTRODUCTION

S-Adenosylmethionine (SAM)-dependent methyltransferases (MTs) constitute a broad class of enzymes that find frequent use in primary and secondary metabolism (Ansari et al., 2008; Liscombe et al., 2012). When embedded within Type I polyketide synthases (PKSs), MTs function to catalyze the formation of α-methyl branches on polyketide chains.[2] The action of MTs within PKS modules provides both an alternative route to the α-methyl branched intermediates typically introduced by methylmalonyl-

CoA specific acyltransferase (AT) domains as well as the primary route for generating *gem*-dimethyl functionality (Ishida et al, 2007; Shecnk et al., 2007).

Polyketides comprise a major fraction of all known natural products, and the molecular factories responsible for synthesizing them represent long-standing targets of protein engineering efforts (Williams, 2013). The diversity of polyketides is remarkable, particularly in light of the few types of enzymatic domains and the simple building blocks employed in their assembly. A series of PKS modules, minimally comprised of a ketosynthase (KS) and an acyl carrier protein (ACP) [as well as an acyltransferase (AT) in *cis*-AT PKSs)], orchestrate two-carbon extensions of polyketide chains *via* decarboxylative condensations. Each module typically contains one or more processing domains, such as a ketoreductase (KR), dehydratase (DH), enoylreductase (ER), and C-methyltransferase (MT) that act on the α-/β-carbons of intermediates to confer complexity to the often large, stereo-dense final products (Keatinge-Clay, 2012). Though considerable effort has been dedicated to characterizing the structure and function of PKS biosynthetic pathways, limited work has been performed on embedded MT domains (Winter et al., 2013; Miller et at., 2002). Consequently, their substrate specificities and timing within the catalytic cycle of PKS modules have remained mysterious (Poust et al., 2015).

Two routes for MT-catalyzed methyl incorporation into polyketides are plausible (Figure 2-1) (Poust et al., 2015). In the first and traditionally-accepted route, methylation immediately follows KS-mediated condensation (aided by the decreased pK_a of the β-ketoacylthioester intermediate). Conversely, in the second route, mono- or dimethylation occurs on the ACP-bound malonyl extender unit prior to condensation. Recent experiments on modules containing MTs catalyzing *gem*-dimethylation within the yersiniabactin and epothilone PKS pathways detected the

formation of dimethylmalonyl-ACP *via* tandem MS and showed them to be suitable substrates for the subsequent condensation reaction indicating that the second route may be operative within these pathways (Poust et al., 2015).

RESULTS AND DISCUSSION

To further elucidate the timing of embedded mono- and dimethylating MT domains, we purified and assayed the five monomethylating MTs and the one *gem*-dimethylating MT from the gephyronic acid biosynthetic pathway of *Cystobacter violaceous* (Figure 2-2) (Young et al., 2013). All six of these MTs display activity towards β-ketoacylthioester substrates but not towards equivalent malonyl thioester substrates, providing evidence that the first route is operative for both mono- and dimethylation within the gephyronic acid synthase. To obtain soluble protein for the five monomethylating MT domains in the gephyronic acid synthase, four MTs were cloned and purified as MT+KR didomains (labeled as GphMT1, GphMT2, GphMT3, and GphMT6 for simplicity) and one as an MT+ER+KR tridomain (GphMT4) (Figure 2-3). A sequence alignment of *cis*-AT MTs reveals that MT domains are typically embedded within the KR structural subdomain, and our studies indicate that the inclusion of this region may be necessary for MT expression (Figure 2-4). Despite its location within a non-reducing module, the *gem*-dimethylating GphMT5 harbors a C-terminal region with striking sequence similarity to the KR structural subdomain (Figure 2-4). GphMT5 was assayed both as a discrete domain and in the context of its complete module (GphH). All six excised MT constructs and GphH were expressed in *E. coli* BL21(DE3) and *E. coli* K207-3 strains, respectively, and purified by Ni-NTA (nickel nitrilotriacetate) chromatography.

All monomethylating MTs were observed to catalyze the methylation of 3-oxopentanoyl-*S*-NAC, 1, to afford 2-methyl-3-oxopentanoyl-*S*-NAC, 2, with varying degrees of conversion, ranging from 2-24% after 72 h (Figure 2-5). Notably, after 72 h, no methylation of malonyl-*S*-NAC or malonyl-CoA was observed by any of the five monomethylating MTs. Taken together, these data suggest that all MT-catalyzed monomethylation occurs exclusively via route 1 in the gephyronic acid biosynthetic pathway. Dimethylation of 1 to afford 2,2-dimethyl-3-oxopentanoyl-*S*-NAC, 3, was not detected. With the exception of the MT+ER+KR tridomain, GphMT4, the KRs in each of the didomain constructs were capable of reducing 1 and 2 (Figures 2-6 and 2-7).

Both the GphH module and the MT excised from it, GphMT5, are capable of dimethylating 1 to afford 3, albeit with low efficiency (Figure 2-8). Slightly higher conversion was observed for the GphH module, implying that the *gem*-dimethylating MT is more active in the context of its surrounding domains. Notably, 2 was not detected from reactions with either GphH or GphMT5, suggesting that the second methylation event occurs quickly after the first. Neither GphH nor GphMT5 catalyzed a detectable amount of methylation of malonyl-*S*-NAC or malonyl-CoA, suggesting that dimethylation, like monomethylation, occurs exclusively via route 1 in the gephyronic acid PKS pathway.

To remove the generated *S*-adenosylhomocysteine (SAH) product, a potential inhibitor of MTs, we assayed the activity of each of the GphMTs in the presence of the *E. coli* SAH nucleosidase Pfs (Liscombe et al., 2012; Hendricks et al., 2004). Co-incubation of GphMTs with Pfs greatly increased conversion of 1 to 2 by GphMT1 and GphMT6, with 95% and 85% conversion, respectively, after 72 h (Figure 2-9). Scaled-up reactions with GphMT1 provided milligram quantities of

2 with isolated yields of 78%. The inclusion of Pfs in GphH and GphMT5 dimethylating reactions did not provide a significant increase in conversion, suggesting that the concentration of SAH generated from the lower conversion of substrate is at sub-inhibitory levels (Figure 2-10).

To gain insight into the substrate tolerance of embedded PKS MT domains, we next assayed the most active domains, GphMT1 and GphMT6, with 2 additional substrates of varying chain length, 3-oxohexanoyl-*S*-NAC, 4, and 3-oxobutanoyl-*S*-NAC, 6 (Figure 2-11). Since 1 most closely mimics the natural substrate of GphMT1, the marked decrease in activity observed for both the larger substrate 4 and the smaller substrate 6 suggests that GphMT1 exerts a moderate degree of selectivity for molecules most closely resembling the natively encountered substrate. Similarly, the decreased conversion of 6 by GphMT6 relative to the larger substrates 1 and 4 suggests that mild selectivity is again arising from protein-substrate interactions beyond the β-keto group. The preference of GphMT6 for larger hydrophobic substrates can be rationalized by its presence within the penultimate module. In aggregate, these results are similar to the recent kinetic characterization of the MT domain from the lovastatin synthase LovB in that despite showing a preference for compounds closely resembling the natural substrates, they are promiscuous enough to act on a variety of β-ketoacyl compounds (Cacho et a., 2015).

It is conceivable that the methylation of small molecule substrate mimics may not accurately reflect the methylation of the natural substrates of MTs. Therefore, we examined the activities of two MTs towards their cognate holo-ACPs bound either to a β-ketopentanoyl moiety or a malonyl extender unit. GphACP1 and GphACP6 (the cognate ACPs of GphMT1 and GphMT6,

respectively) were expressed in *E. coli* K207-3 to afford *holo*-ACP domains that were then thioesterified with either substrate 1 or malonyl-CoA. For both GphMT constructs a 14-Da mass increase was observed by HRMS analysis (Figure 2-12). MS/MS spectra confirm MT-catalyzed methylation occurred at the α-position of the diketide attached to the pantetheinyl arm on the conserved ACP serine. Consistent with the previous results, no mass increase was observed during methylation attempts of either of the malonyl-*S*-ACPs. The selective methylation of β-ketoacyl-*S*-ACPs provides further evidence that MT-catalyzed B-branching occurs via route 1 during gephyronic acid biosynthesis.

CONCLUSION

In closing, our results provide significant evidence that MT-catalyzed methylation immediately follows KS-mediated condensation in the gephyronic acid PKS pathway. The data also represents the first reported activity of mono- and dimethylating MTs excised from a modular PKS and helps to provide insight into the substrate promiscuity of embedded MTs as well as the order of chemistry performed by PKS processing domains. As it stands, the reported variability in the timing of α-branching presents MTs as unique, versatile domains within PKS pathways that requires further exploration (Poust, et al., 2015; Young et al., 2013). Additionally, the robust activity of several of the GphMT domains when coupled with Pfs activity, make them potential tools for the incorporation of α-branches into previously reported PKS biocatalytic platforms to allow a greater access to polyketide chemical diversity.

ACKNOWLEDGEMENTS

The following funded this work: The National Institutes of Health (GM106112), the National Science Foundation (CHE1402753), and the Welch Foundation (F-1712,

AKC; F-1155, JSB). We would like to thank the University of Texas at Austin mass spectrometry facility for their help in obtaining high-resolution masses for **1**-**7**.

METHODS AND MATERIALS

Cloning and Expression of MT Domains

The DNA encoding all polypeptides was amplified using primers listed in table 2-1 from *Cystobacter violaceus* or *E. coli* BL21(DE3) genomic DNA and inserted into pGAY28b, a ligation independent cloning vector constructed from pET28b (Gay et al., 2014). *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} = 0.4, and induced with 0.5 mM IPTG. After 18 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (0.5 M NaCl, 10% (v/v) glycerol, 0.1 M HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 x g, 30 min). The supernatant was poured over a column of nickel-NTA resin (Thermoscientific), which was then washed with 40 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 150 mM imidazole. The eluted protein was concentrated to ~10 mg/mL in the equilibration buffer and stored at -80 ˚C until needed.

Substrate Synthesis

3-oxopentanoyl-*S*-NAC (**1**), 2-methyl-3-oxopentanoyl-*S*-NAC (**2**), 3-oxohexanoyl-*S*-NAC (**4**), and 3-oxobutanoyl-*S*-NAC (**6**) were synthesized as previously reported.[2]

In Vitro MT Reaction Conditions

Reaction conditions for 200 μL monomethylating MT reactions: either **1** (10 mM), **2** (10 mM) or malonyl-*S*-NAC (10 mM); Tris-HCl (150 mM); NaCl (100 mM); SAM (15 mM); GphMT (50 μM); and glycerol (10% v/v).

Reactions conditions for 200 μL dimethylating MT reactions: either **1** (10 mM), **2** (10 mM) or malonyl-*S*-NAC (10 mM); Tris-HCl (150 mM); NaCl (100 mM); SAM (30 mM); GphMT (50 μM); and glycerol (10% v/v).

Reaction conditions for 200 μL monomethylating MT reactions containing Pfs: either **1** (10 mM), **2** (10 mM) or malonyl-*S*-NAC (10 mM); Tris-HCl (150 mM); NaCl (100 mM); SAM (15 mM); Pfs (25 μM); GphMT (50 μM); and glycerol (10% v/v). High conversion of **1** to **2** was observed at 25-50 μM Pfs; Pfs concentrations below 25 μM provided lower conversions.

Reactions conditions for 200 μL dimethylating MT reactions containing Pfs were as follows, either **1** (10 mM), **2** (10 mM) or malonyl-*S*-NAC (10 mM); Tris-HCl (150 mM); NaCl (100 mM); SAM (30 mM); Pfs (25 μM); GphMT (50 μM); and glycerol (10% v/v).

Reactions were run at ~25 °C for 24-72 h. After completion, reactions with diketide **1** were extracted with 3 reaction volumes of EtOAc. The organic layer was dried *in vacuo* and resuspended in MeOH for HPLC analysis. After completion, reactions with malonyl-*S*-NAC were heated to precipitate enzyme and centrifuged to clarify the supernatant for HPLC analysis.

Scaled 20 mL GphMT1 reactions: **1** (10 mM), Tris-HCl (150 mM); NaCl (100 mM); SAM (30 mM); Pfs (50 μM); GphMT (50 μM); and glycerol (10% v/v). Reaction was stirred at room temperature for 60 h and extracted with 3 volumes of EtOAc (3 x 20 mL). The EtOAc layer was washed with brine, dried over $MgSO_4$, and concentrated *in vacuo* to provide a yellow oil that was purified *via* flash chromatography (75% EtOAc:hexanes) to provide **2** as a yellow oil (36 mg, 78%).

Methylation of excised acyl-ACPs: Thiol reactions to produce acyl-ACPs were **1** (5 mM), $NaHCO_3$ pH 8.1(150mM), and GphACP(1/6) (50 μM). Thiol reactions to produce malonyl-ACPs were malonyl-CoA (5 mM), $NaHCO_3$ pH 8.1 (300mM), and GphACP(1/6) (50 μM). Thiol exchange were run at room temperature for ~30m. Following thiol exchange, Acyl-/malonyl-ACPs were collected with a 10 kDa spin filter. Methylation reactions to monitor methylation of excised ACPs were malonyl-/acyl-ACP (~10 μM), Tris-HCl (150 mM); NaCl (100 mM); SAM (10 mM); Pfs (20 μM); GphMT (20 μM) at room temperature for ~16h. Methylation reactions were first filtered with a 30kDa spin filter to remove GphMTs and Pfs followed by a 10 kDa spin filter to collect malonyl-/acyl-ACPs for MS analysis.

HPLC and HRMS Analysis

HPLC conditions: All HPLC monitoring was performed on a tandem Waters 2707 autosampler and Waters 1525 binary HPLC pump connected to a Waters 2998 photodiode array detector using a Varian Microsorb-MV C18 column (250 x 4.6 mm, 5 μm particle size, 100 Å pore size) and mobile phases consisting of water with 0.1% TFA

(solvent A) and acetonitrile with 0.1% TFA (solvent B) with a solvent gradient of 5%-100% B over 30 min at a flow rate of 1 mL/min. HRMS conditions: High-resolution mass spectrometry measurements were obtained by chemical ionization (ESI) with a VG analytical ZAB2-E instrument. HRMS & MS/MS conditions: All experiments were performed on a Thermo Fisher Scientific Orbitrap Elite mass spectrometer (San Jose, CA) modified with a 193 nm ArF excimer laser (Coherent ExciStar XS) to perform ultraviolet photodissociation (UVPD) experiments in the HCD cell. Protein solutions were loaded into Au-coated borosilicate emitters (10 μM in 50% (v/v) acetonitrile) and ionized offline using a nano-ESI source or by separation using a Dionex UltiMate 3000 RPLC nanoLC system integrated with an ESI source. Chromatographic separations were performed using water (mobile phase A) and acetonitrile (mobile phase B), each containing 0.1% formic acid with a solvent gradient of 15% to 55% B over 45 min at a flow rate of 300 nL/min. The trap (30 mm × 0.1 mm) and analytical columns (with integrated emitter) (20 cm × 0.075 cm) were packed in-house using 5 μm Agilent PLRP-S resin (1000 Å pore size). CID mass spectra were collected using normalized collision energy 25%. UV photoactivation of the modified proteins was achieved using one pulse with energy 2 mJ/pulse. MS1 and MS/MS spectra (50-250 scans) were acquired at 240 K resolving power (at m/z 400). Processing of MS/MS spectra was performed using Xtract and ProSight 3.0 to identify fragment ions. 1H-NMR, HRMS and HPLC of MT reactions below (Figures 2-13 – 2-41)

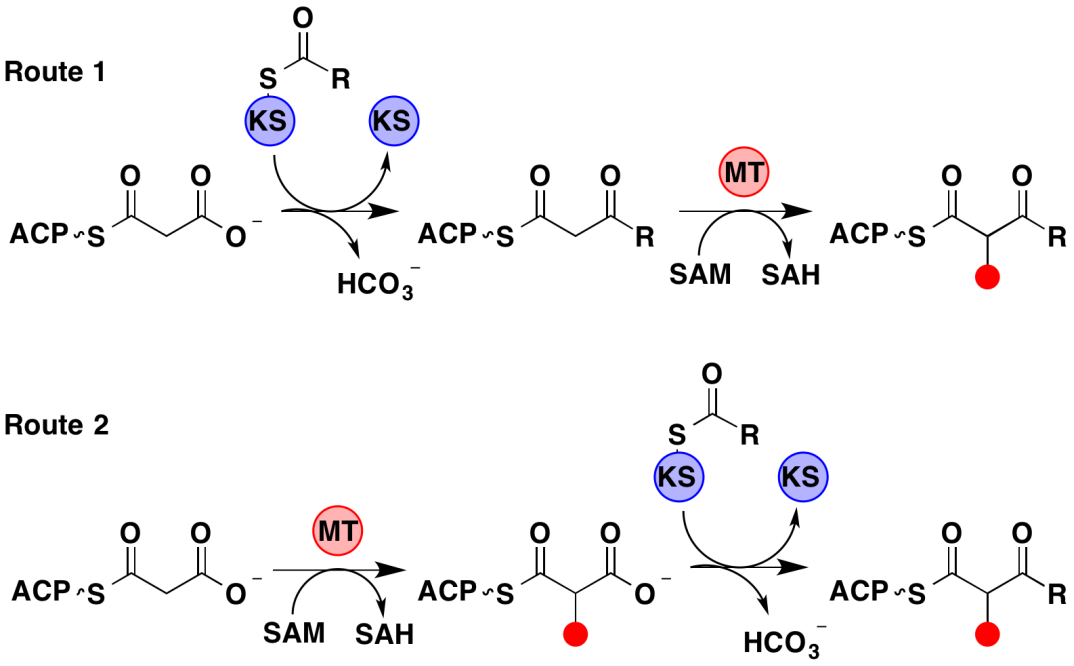


Figure 2-1: Timing of methyltransferase domains.

Potential routes for polyketide synthase methyltransferase-catalyzed α-methylation. In route 1 methylation occurs immediately after ketosynthase mediated condensation. In route 2 methylation precedes condensation using a malonyl-ACP extender unit as substrate.

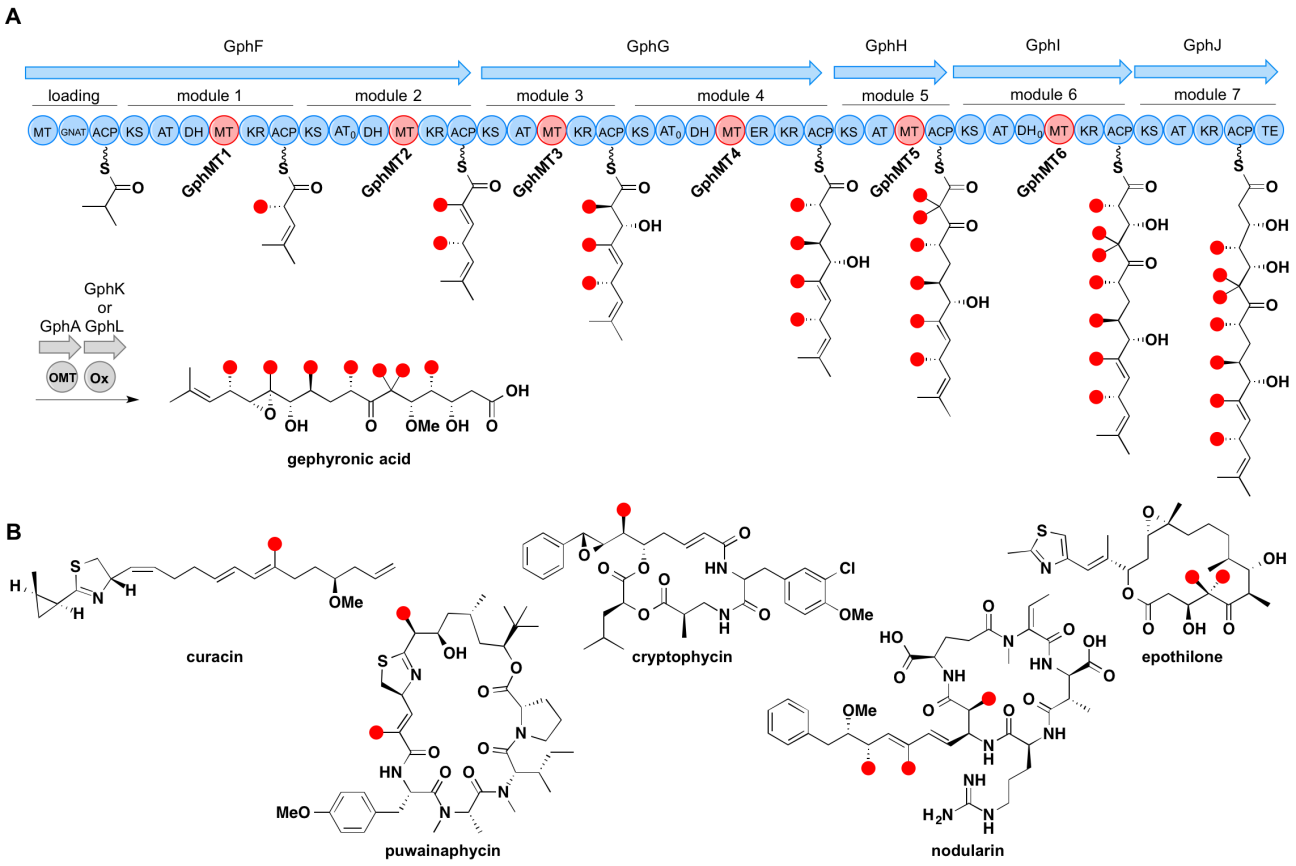


Figure 2-2: Methytransferases in gephyronic biosynthesis.

(A) Model of the gephyronic acid synthase and structure of gephyronic acid. MT domains highlighted in red. (B) Polyketides containing α-methyl branches (red) installed by MT domains embedded within *cis*-acyltransferase PKSs.

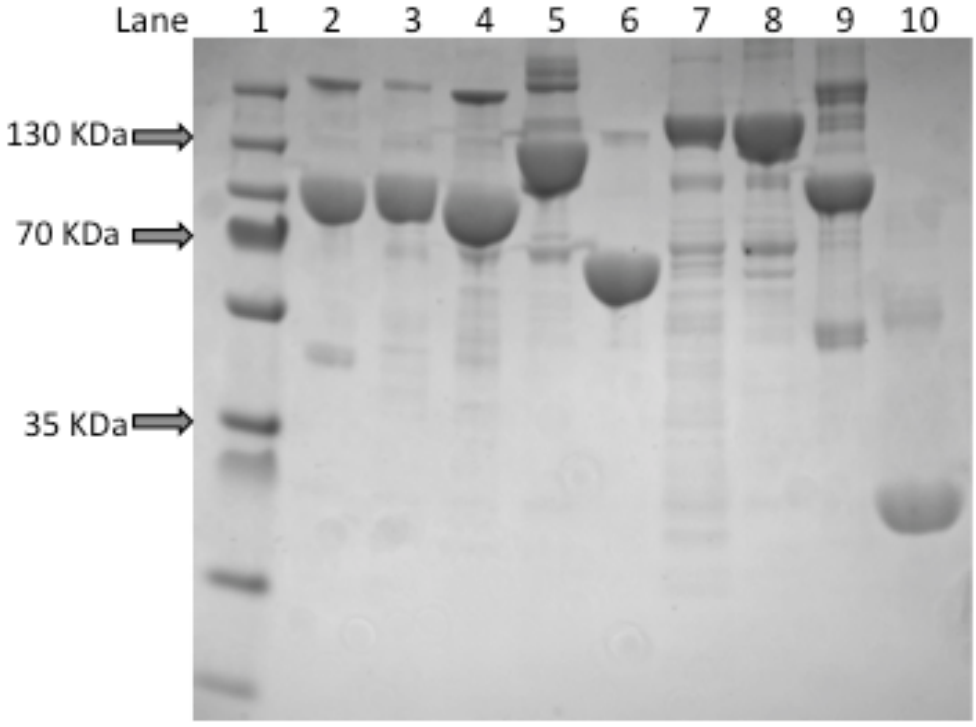


Figure 2-3: SDS-PAGE of purified MT-containing PKS fragments.

1) MW marker, 2) GphMT1, 3) GphMT2, 4) GphMT3, 5) GphMT4, 6) GphMT5, 7) GphH (phosphopantetheinylated), 8) GphH, 9) GphMT6, 10) Pfs.

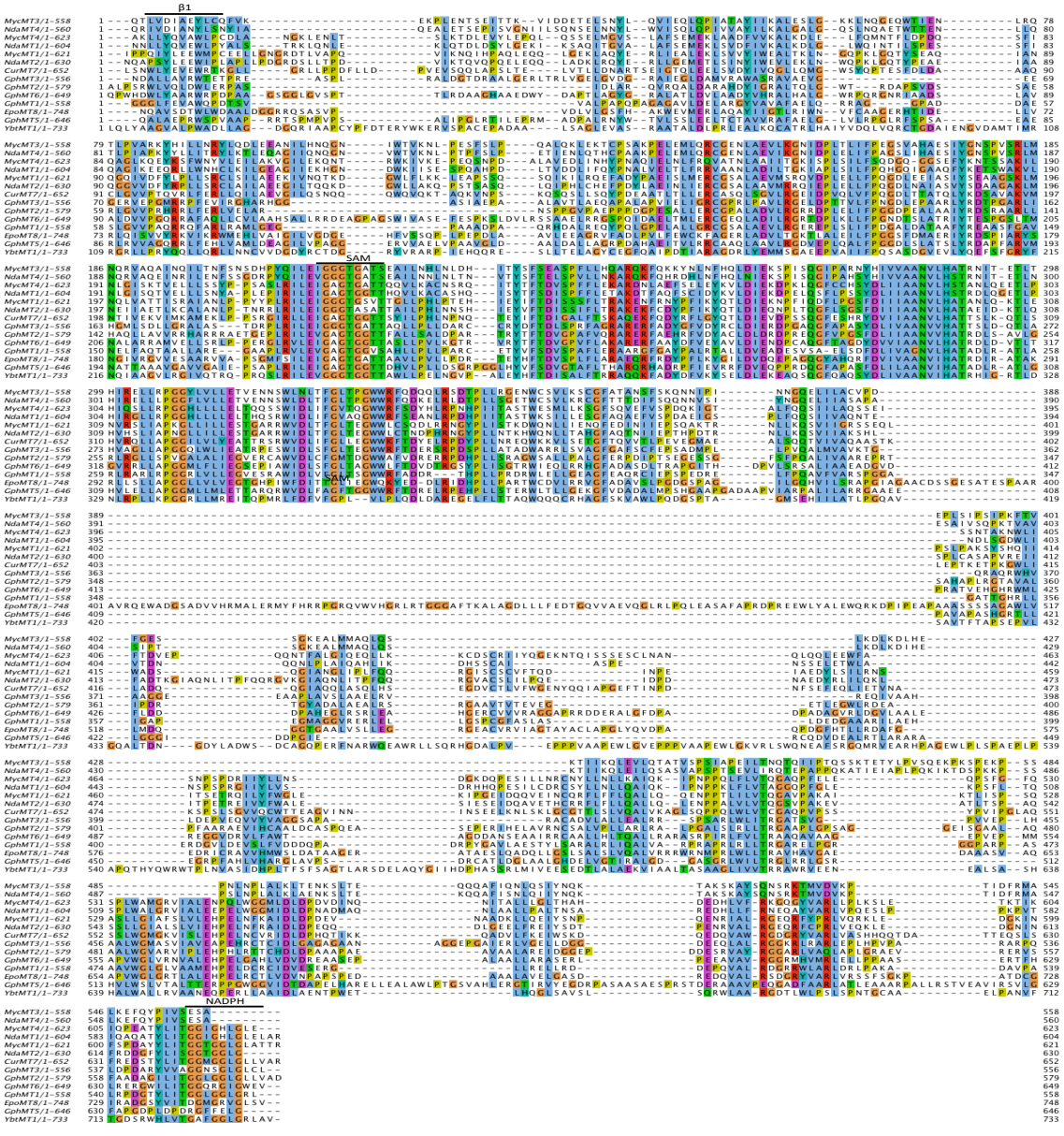


Figure 2-4. Multiple sequence alignment of *cis*-AT MTs.

Myc = microcystin, Nda = nodularin, Gph = gephyronic acid, Cur = curacin, Epo = epothilone, and Ybt = yersiniabactin. 'B1' indicates first beta-strand of KR structural subdomain. The two subsequent black lines indicate SAM binding motif of MT domain and NADPH binding motif that begins the KR catalytic subdomain, respectively

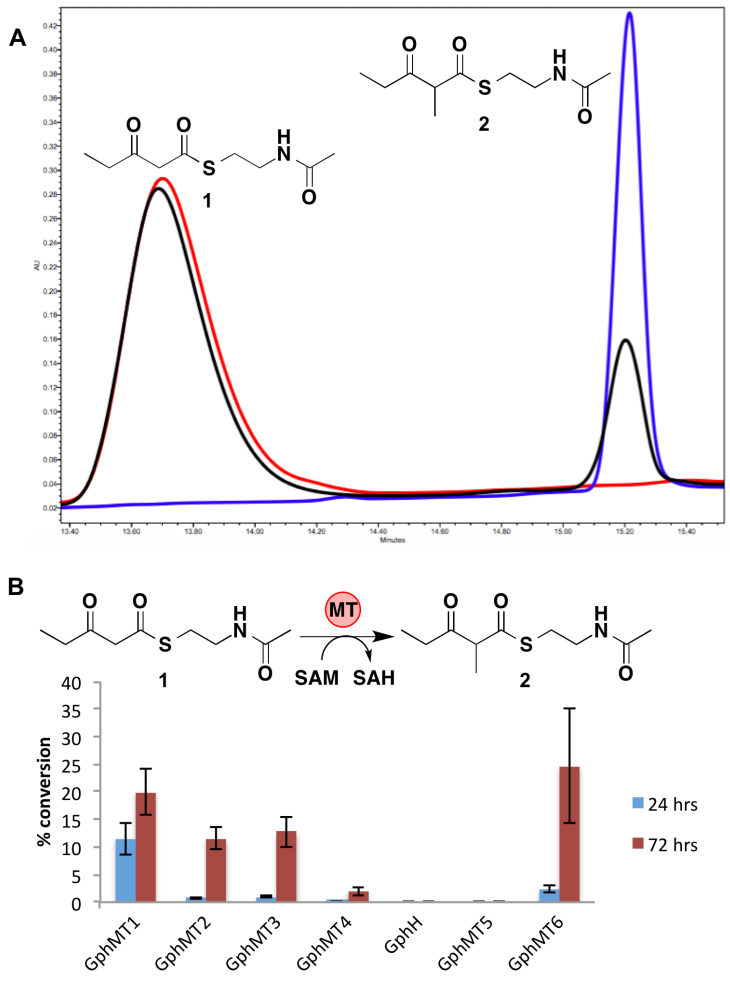


Figure 2-5: GphMT activity.

(A) HPLC traces monitoring 235 nm; red, 1 synthetic standard; blue, 2 synthetic standard; black, GphMT1 reaction after 72 hrs. All HPLC peaks collected and confirmed via HRMS. (B) Conversion of 1 to 2 catalyzed by 50 uM MT constructs and GphH with 10 mM 1 and 15 mM SAM. Error bars represent standard deviation calculated from triplicate data.

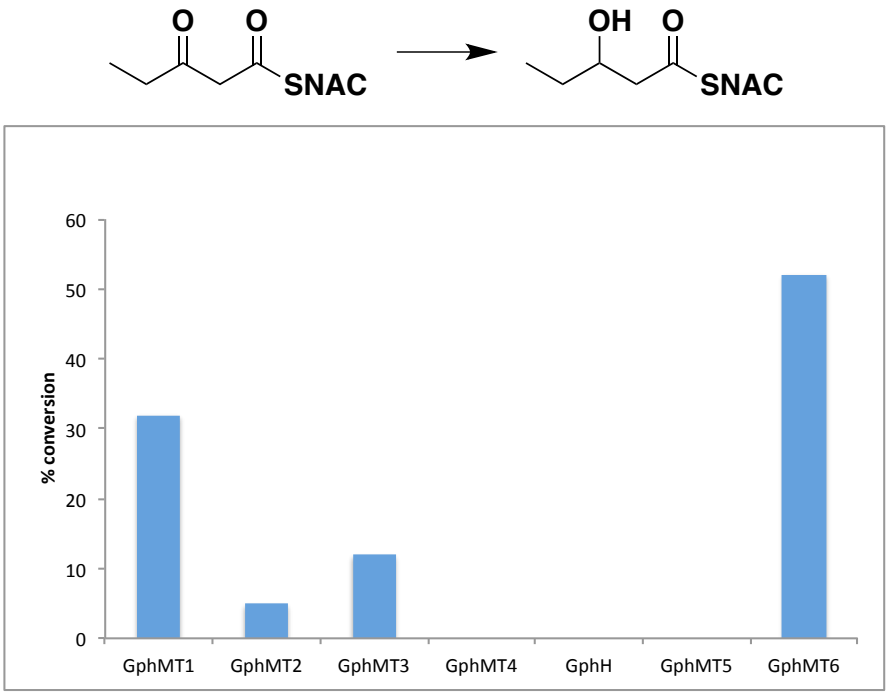


Figure 2-6: Gph KR activity.

Reduction of 1 by GphMT constructs with relative conversions. *In vitro* reductions were run as previously reported.

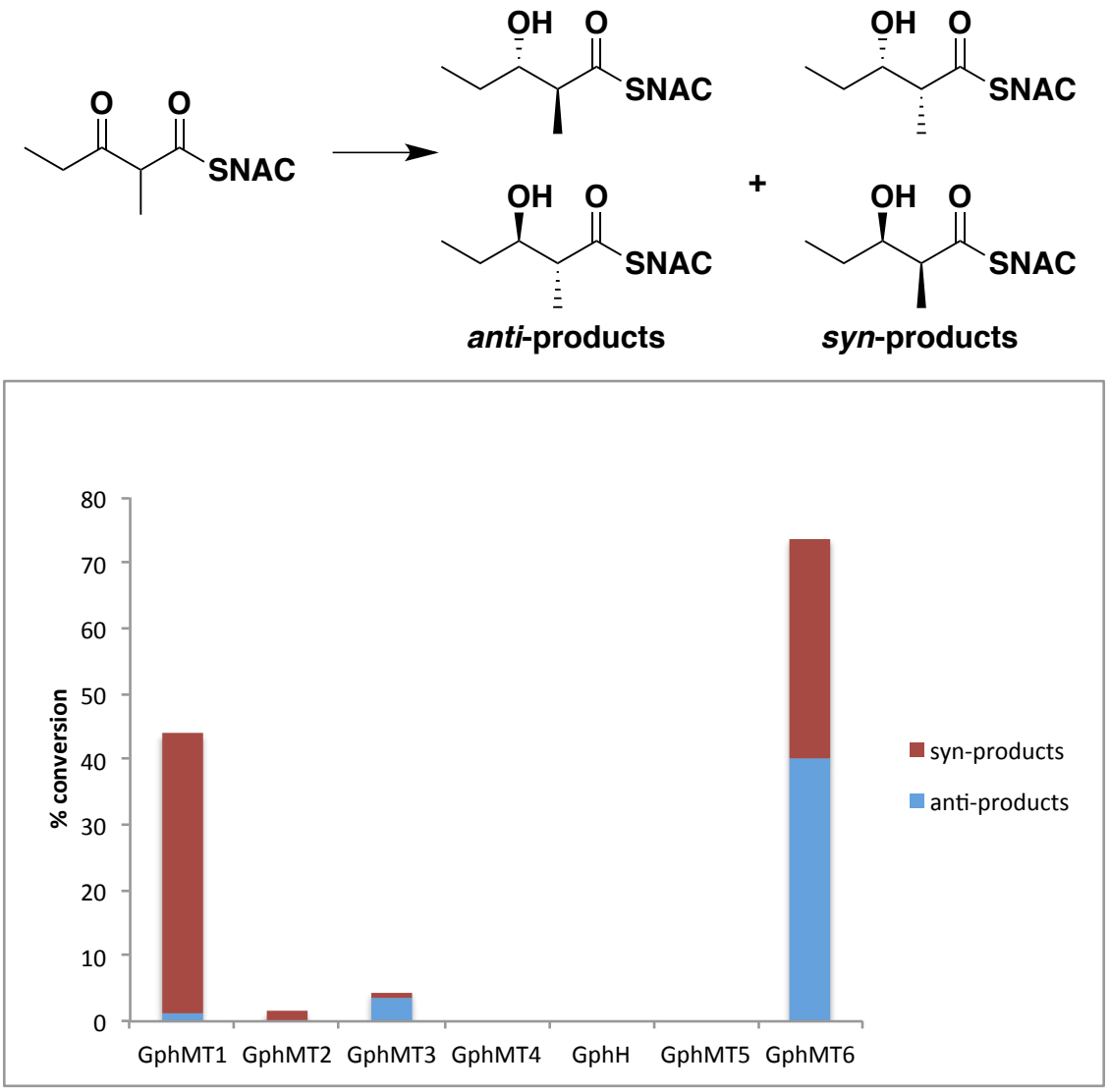


Figure 2-7: Reduction of by KR-containing GphMT constructs.

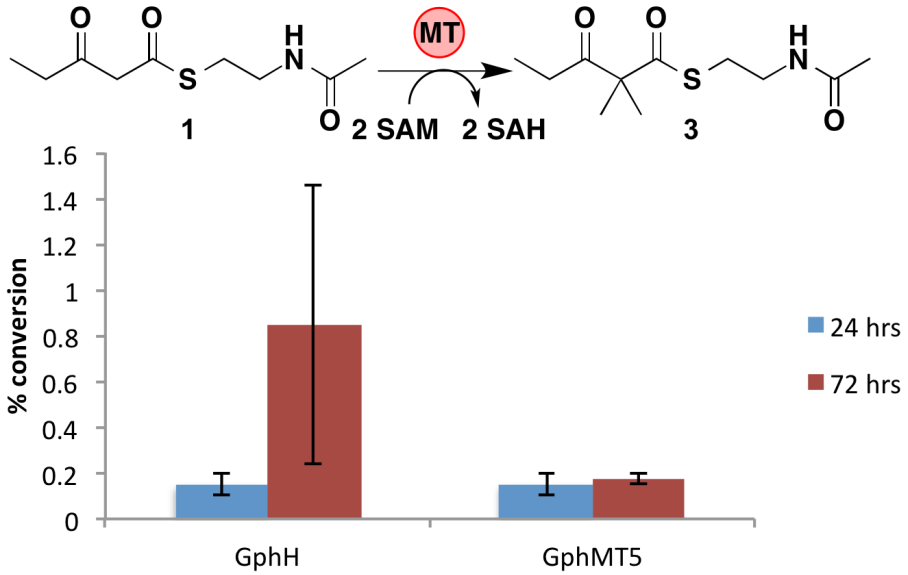


Figure 2-8: Dimethylation activity of GphMT5.

Conversion of 1 to 3 catalyzed by 50 uM GphH and GphMT5 with 10 mM 1 and 30 mM SAM. Error bars represent standard deviation calculated from triplicate data.

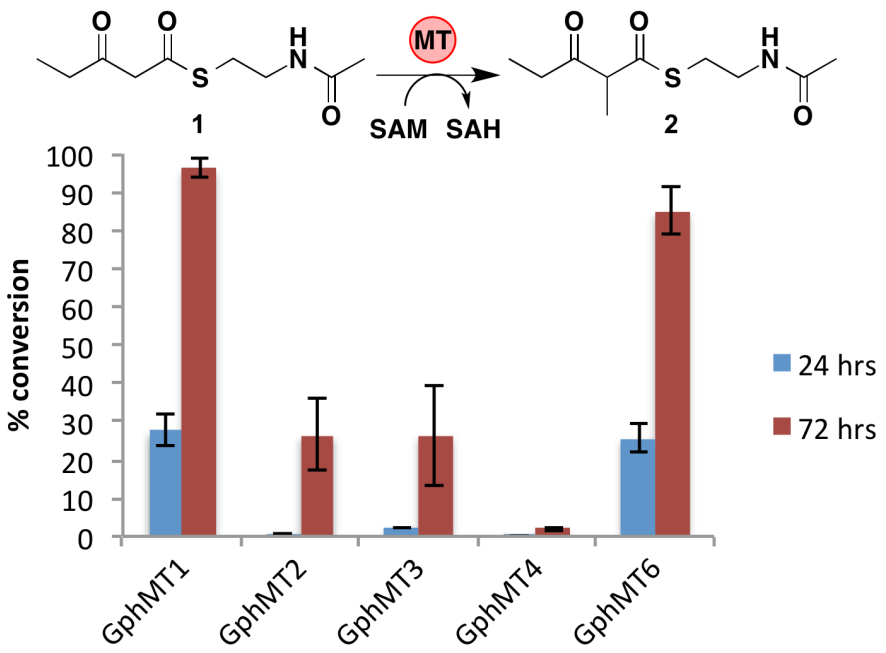


Figure 2-9: GphMT activity with Pfs.

Conversion of 1 to 2 catalyzed by 50 uM monomethylating GphMT constructs in the presence of 25 uM Pfs with 10 mM 1 and 15 mM SAM. Error bars represent standard deviation calculated from triplicate data.

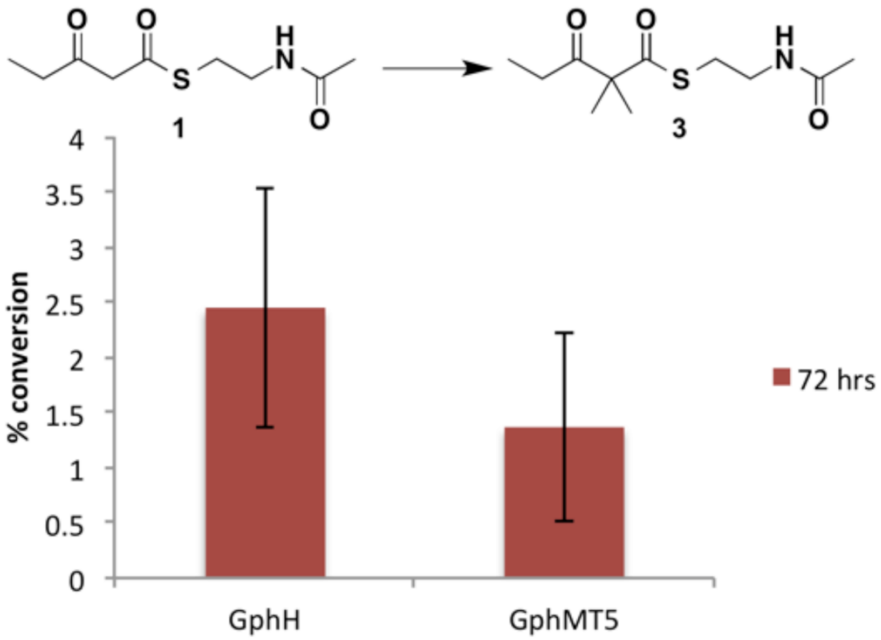


Figure 2-10: GphMT5 activity with Pfs.

Conversion of 1 to 3 via GphH and MT5 catalyzed dimethylation in the presence of Pfs.

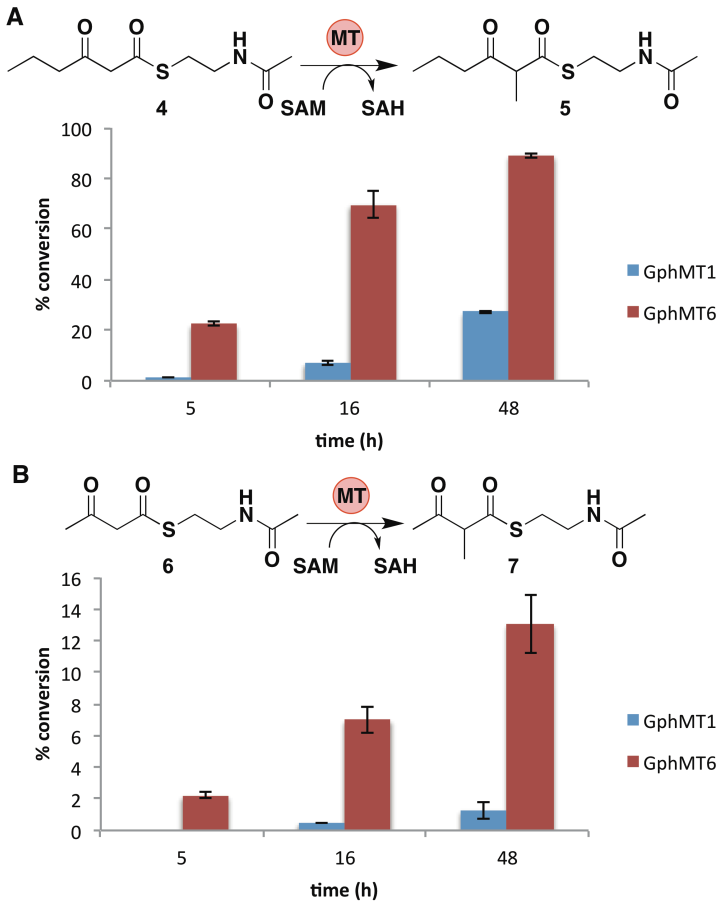


Figure 2-11: Impact of chain length on GphMT1 and GphMT6.

Activity using substrates 4 (A) and 6 (B) (10 mM) with 15 mM SAM, 50 uM MT, and 25 uM Pfs (Note the difference in y-axis scale). Error bars represent standard deviation calculated from triplicate data.

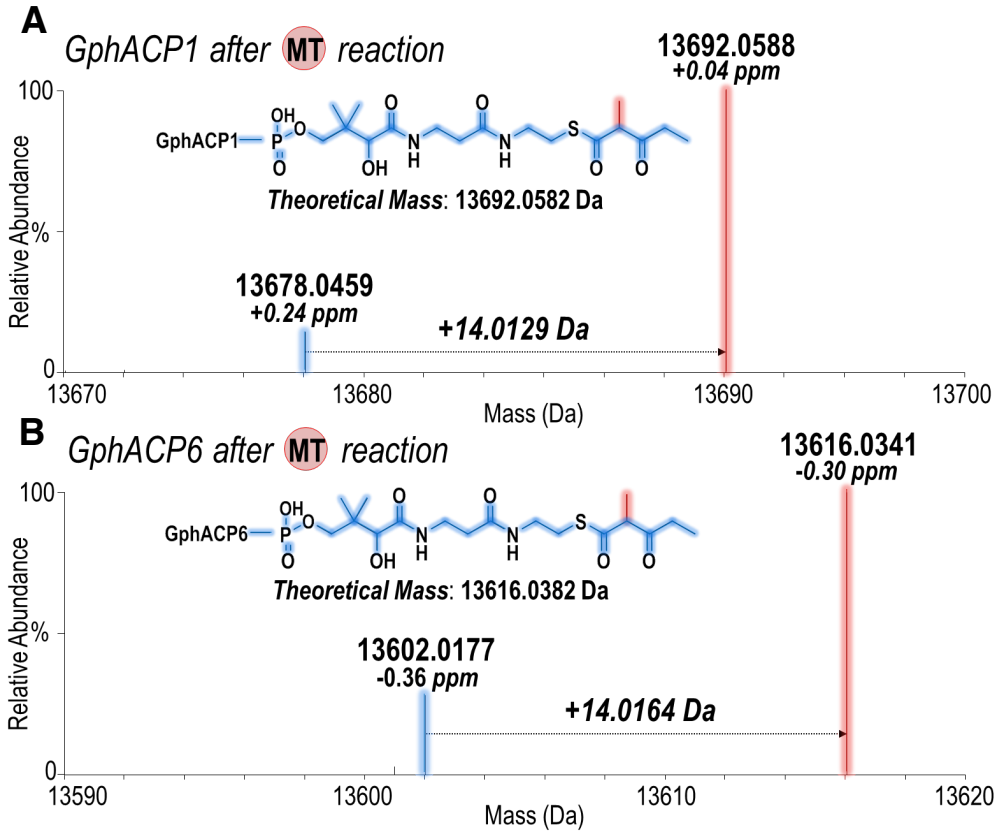


Figure 2-12: ESI-MS of MTs on ACP bound substrates.

Deconvoluted ESI-MS (mass range 13590 – 13620 Da) of phosphopantetheinylated (A) GphACP1 and (B) GphACP6 bound to a 3-ketopentanoyl group showing a 14-Da mass increase after the MT reaction.

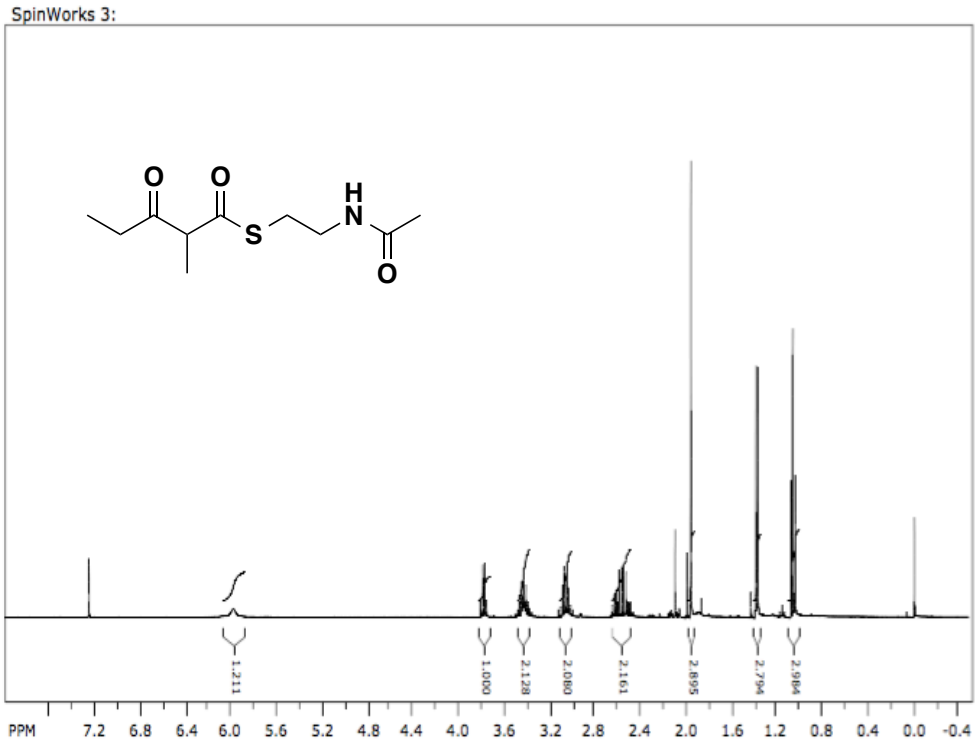


Figure 2-13: ^{1}H-NMR of methylated substrate.

^{1}H-NMR (400 MHz, $CDCl_3$): 6.0 (s, broad, 1H), 3.8 (q, 1H), 3.4 (m, 2H), 3.1 (m, 2H), 2.6 (m, 2H), 1.9 (s, 3H), 1.4 (d, J = 7.2 Hz, 3H), 1.1 (t, J = 7.1 Hz, 3H).

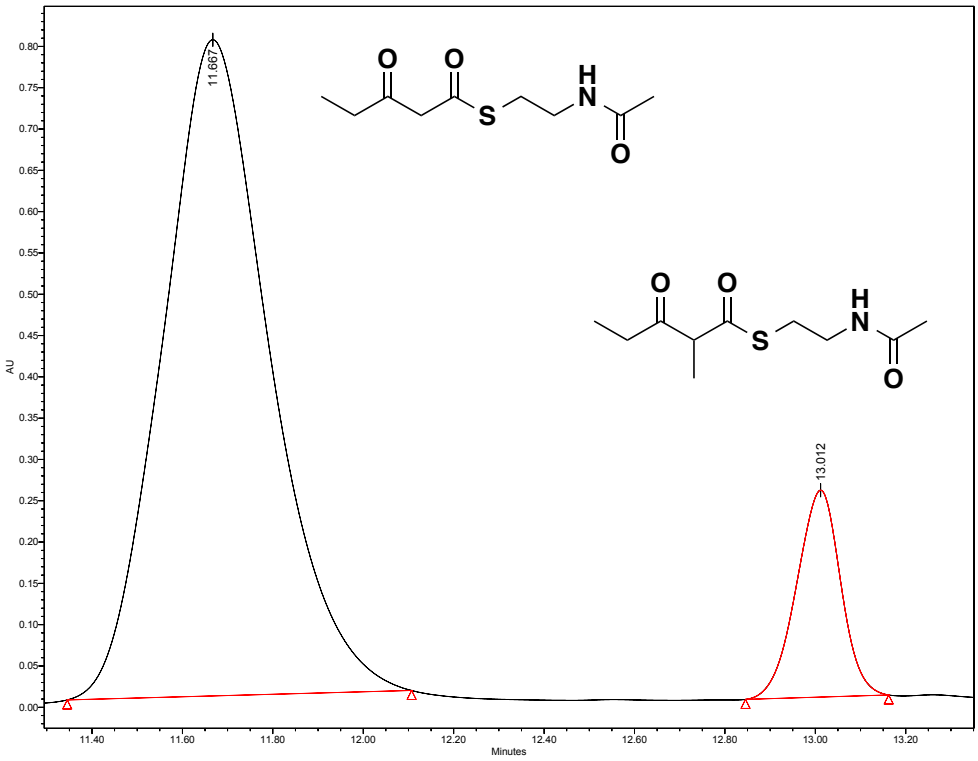


Figure 2-14: GphMT1 reaction HPLC trace 72 h without Pfs.

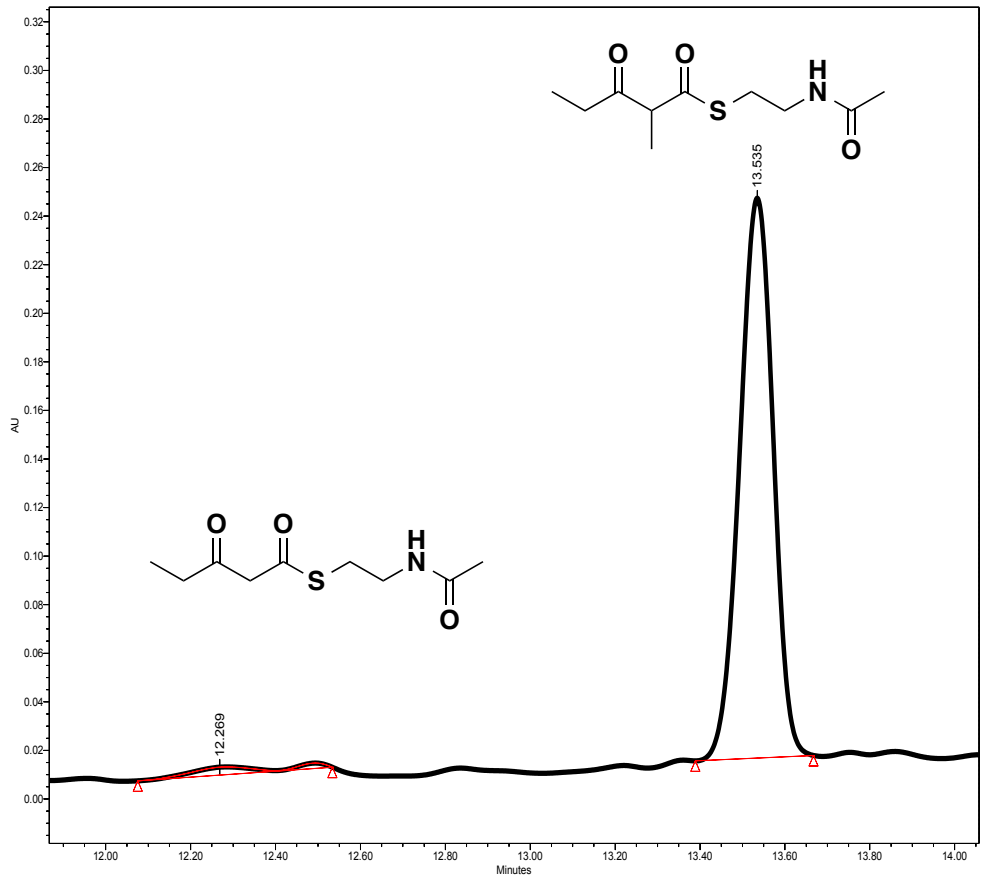


Figure 2-15: GphMT1 reaction HPLC trace 72 h with Pfs.

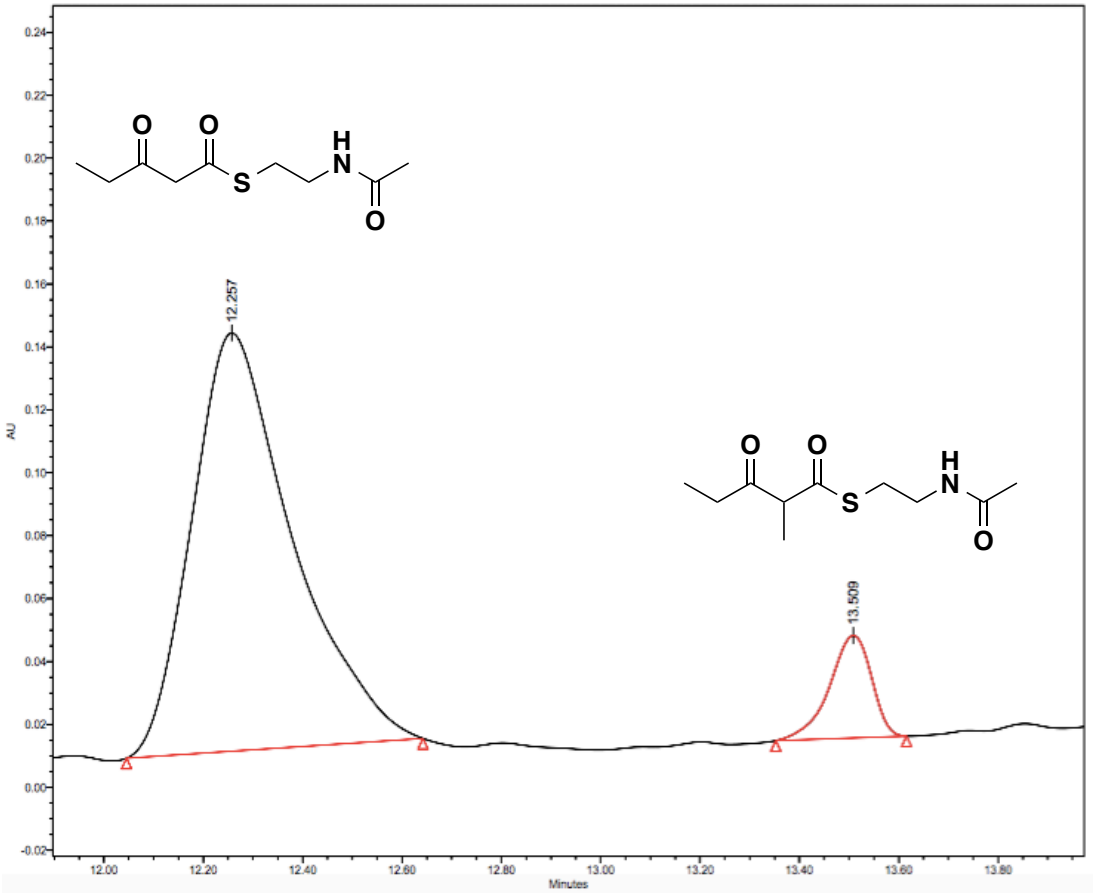


Figure 2-16: GphMT2 reaction HPLC trace 72 h without Pfs.

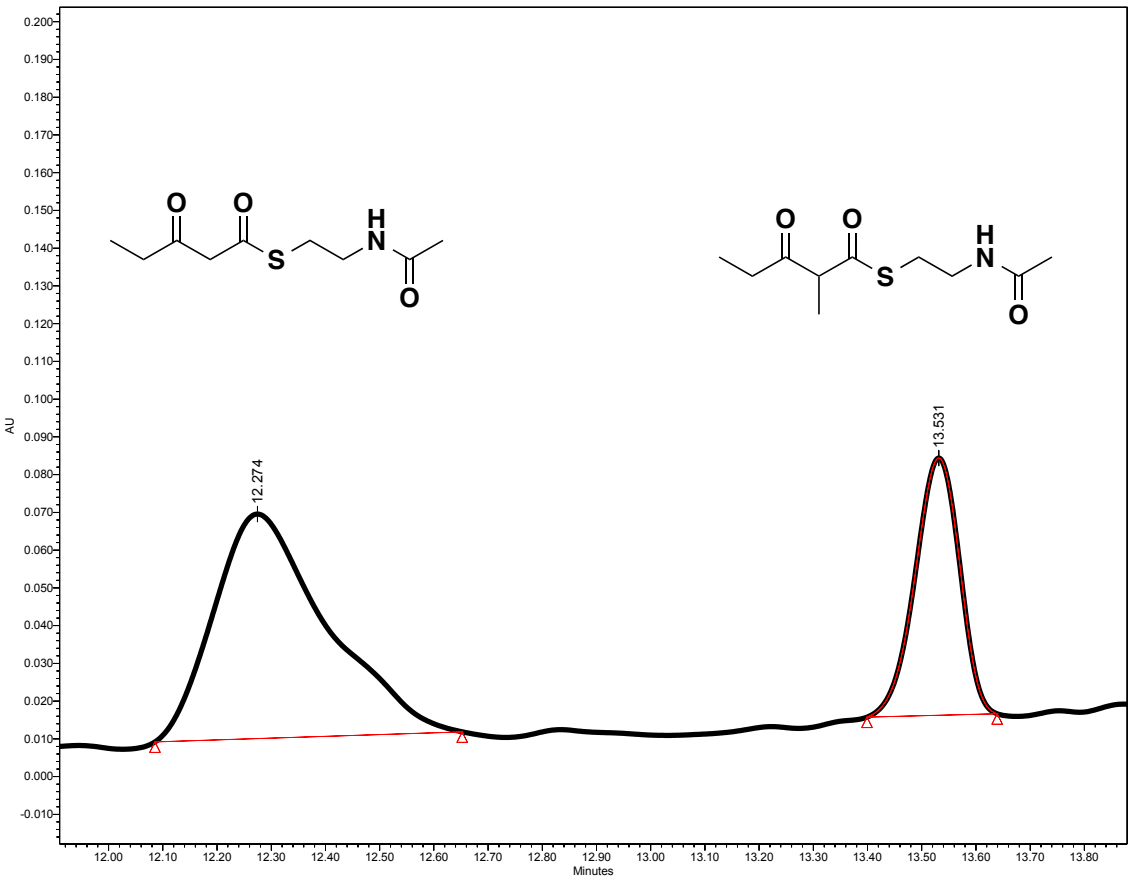


Figure 2-17: GphMT2 reaction HPLC trace 72 h with Pfs.

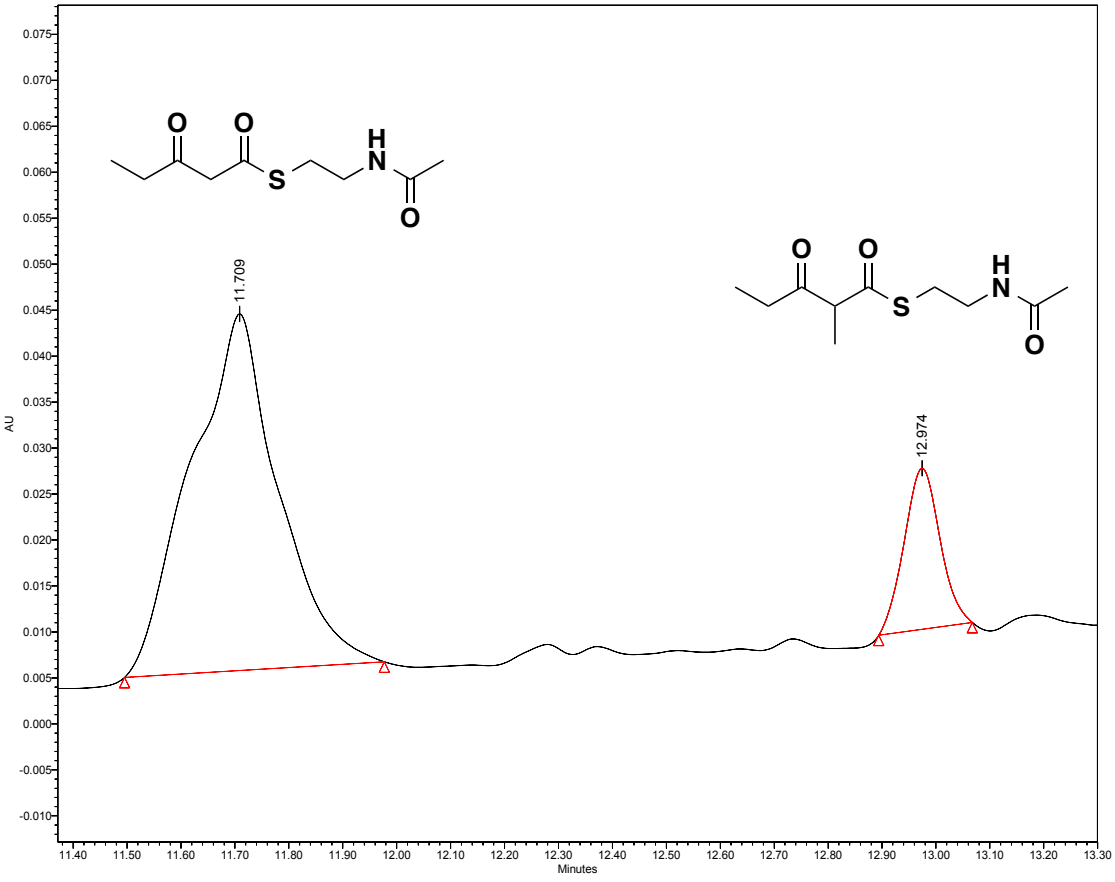


Figure 2-18: GphMT3 reaction HPLC trace 72 h without Pfs.

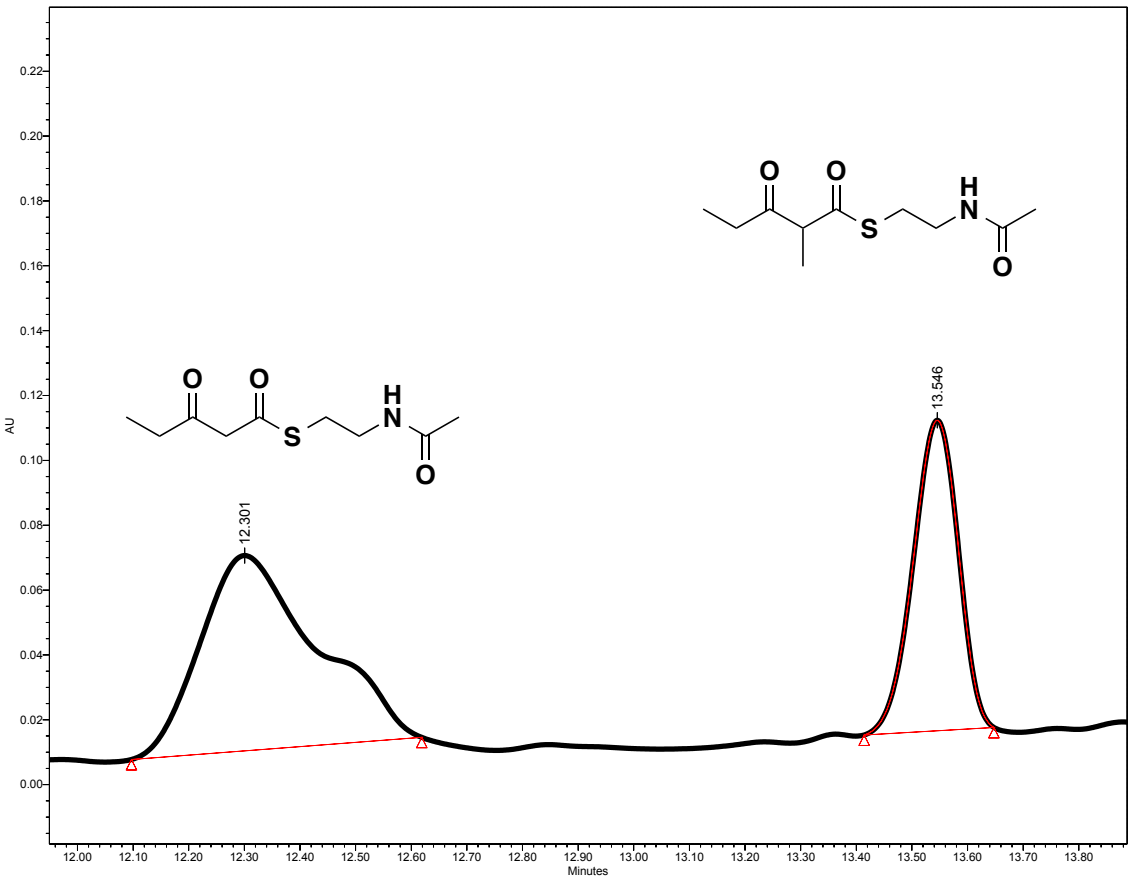


Figure 2-19: GphMT3 reaction HPLC trace 72 h with Pfs.

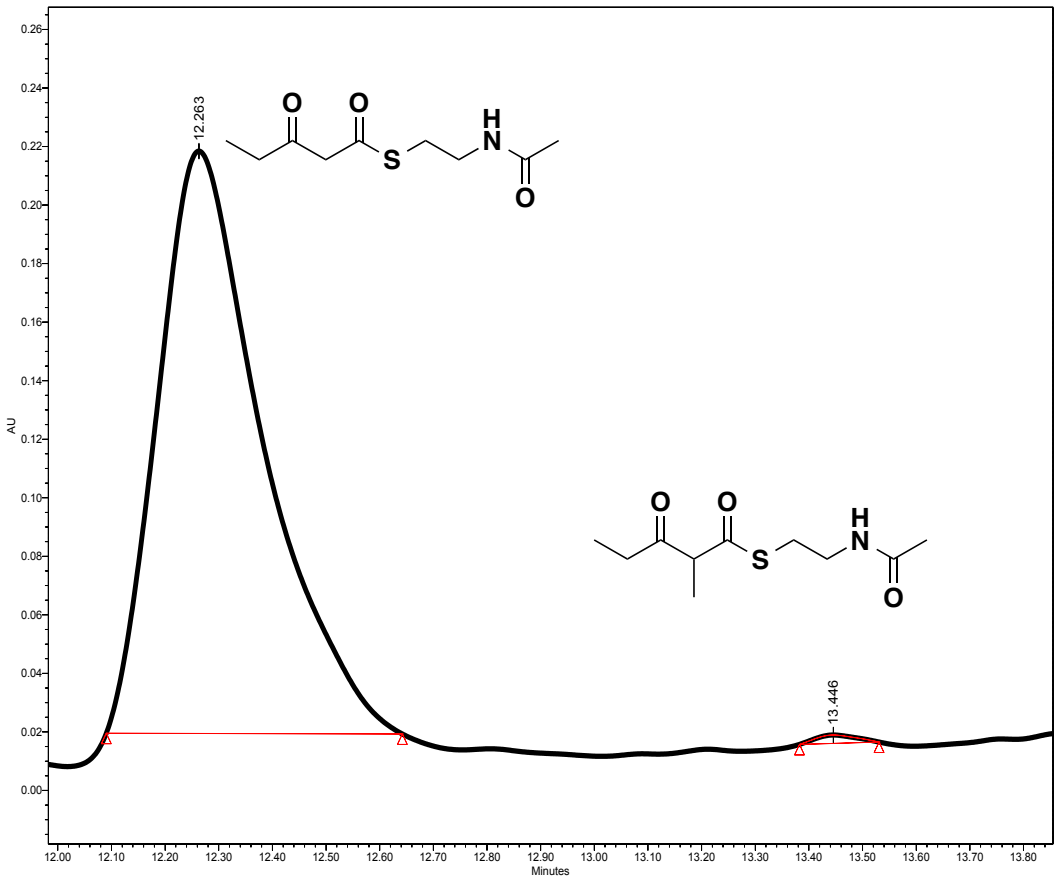


Figure 2-20: GphMT4 reaction HPLC trace 72 h without Pfs.

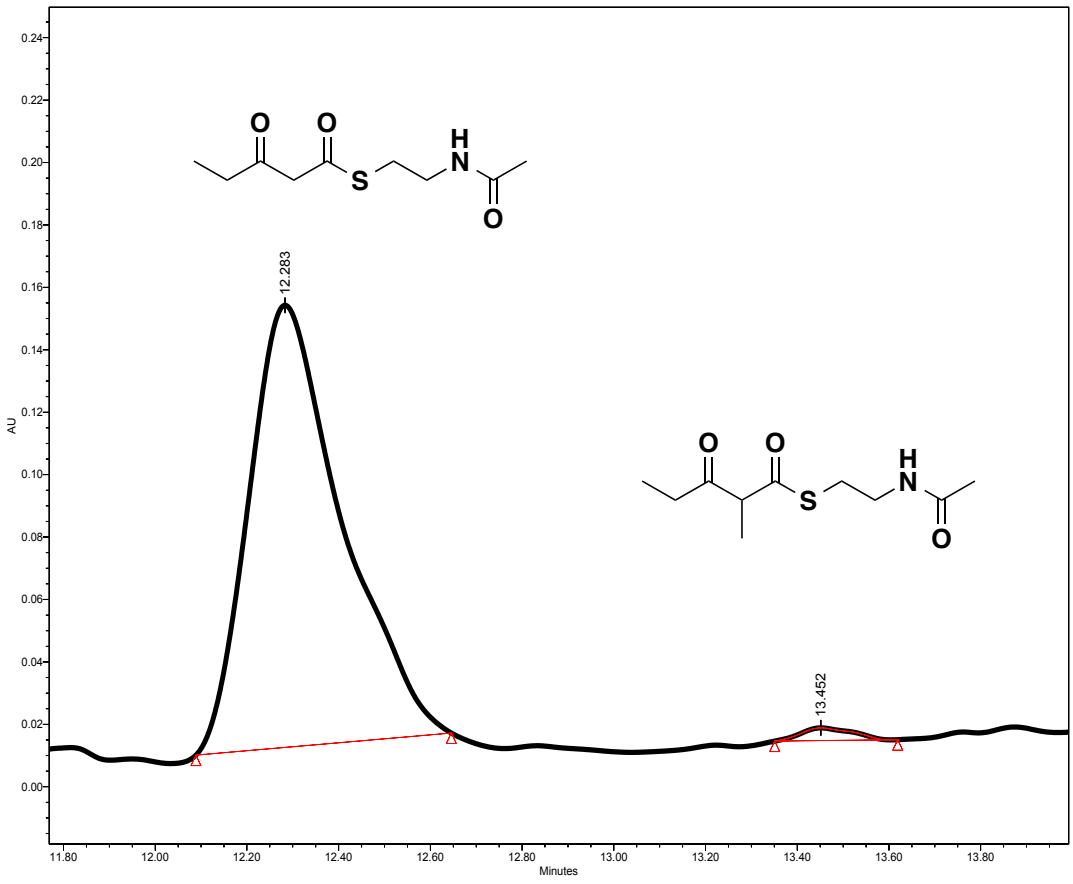


Figure 2-21: GphMT4 reaction HPLC trace 72 h with Pfs.

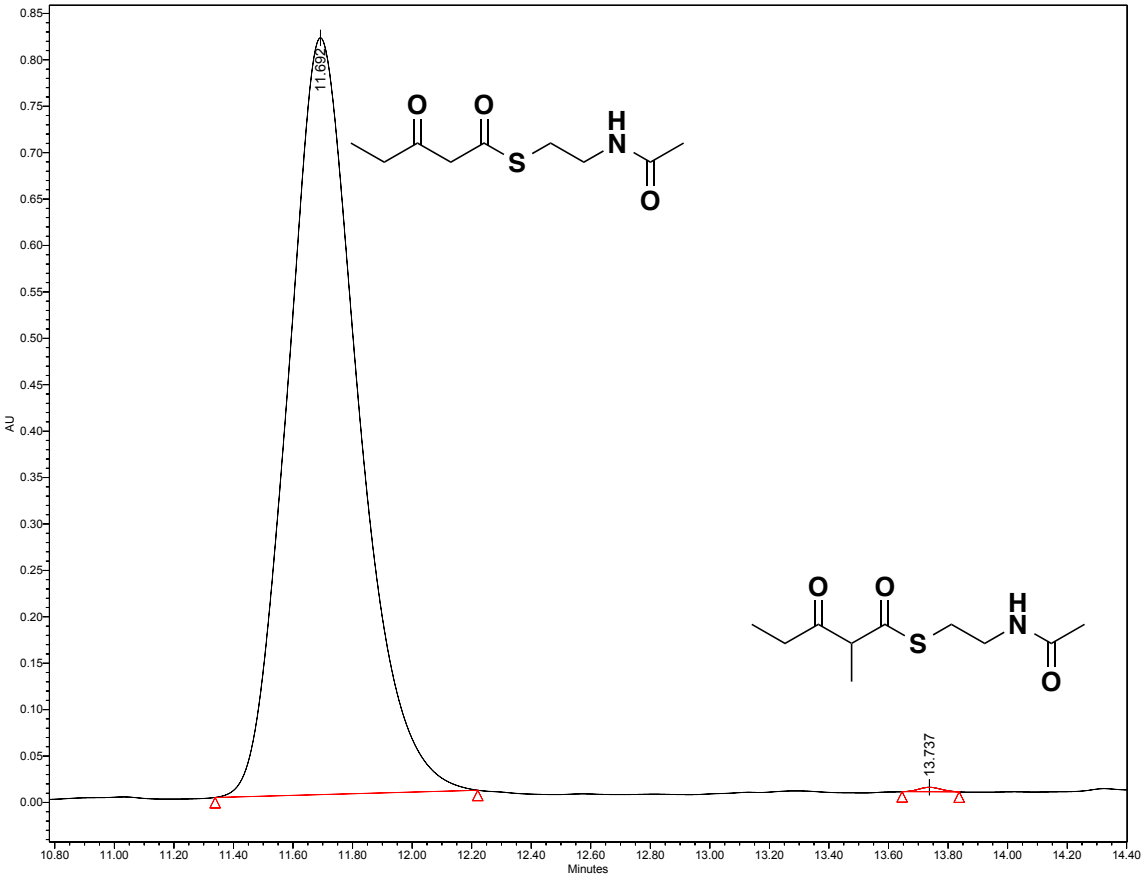


Figure 2-22: GphH module reaction HPLC trace 72 h without Pfs.

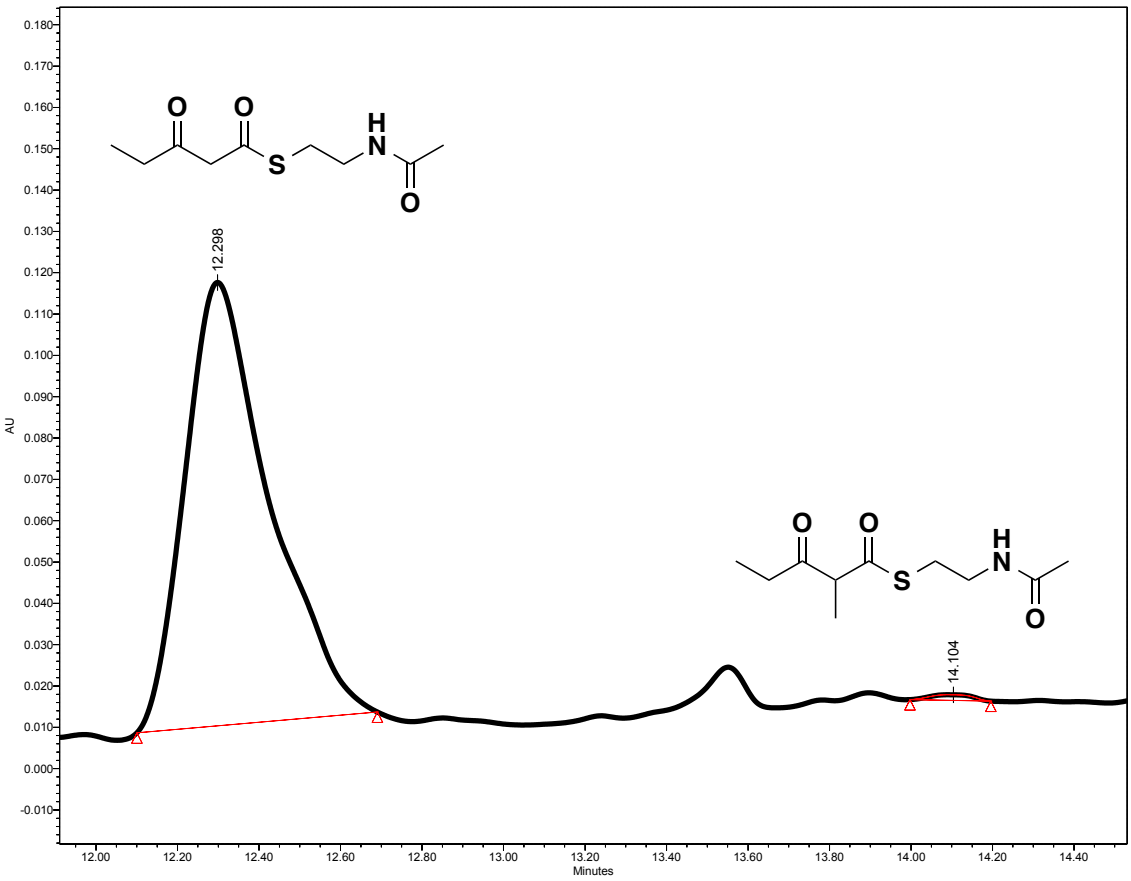


Figure 2-23: GphH module reaction HPLC trace 72 h with Pfs.

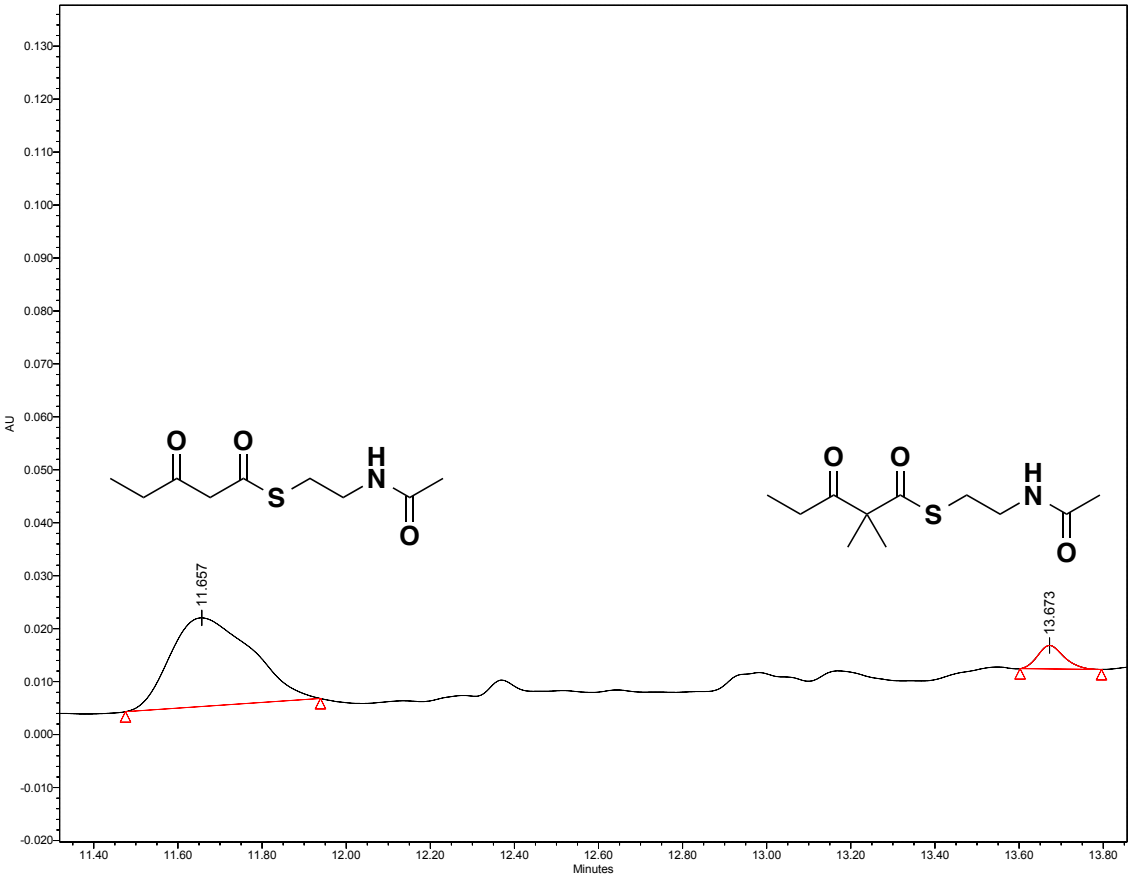


Figure 2-24: GphMT5 reaction HPLC trace 72 h without Pfs.

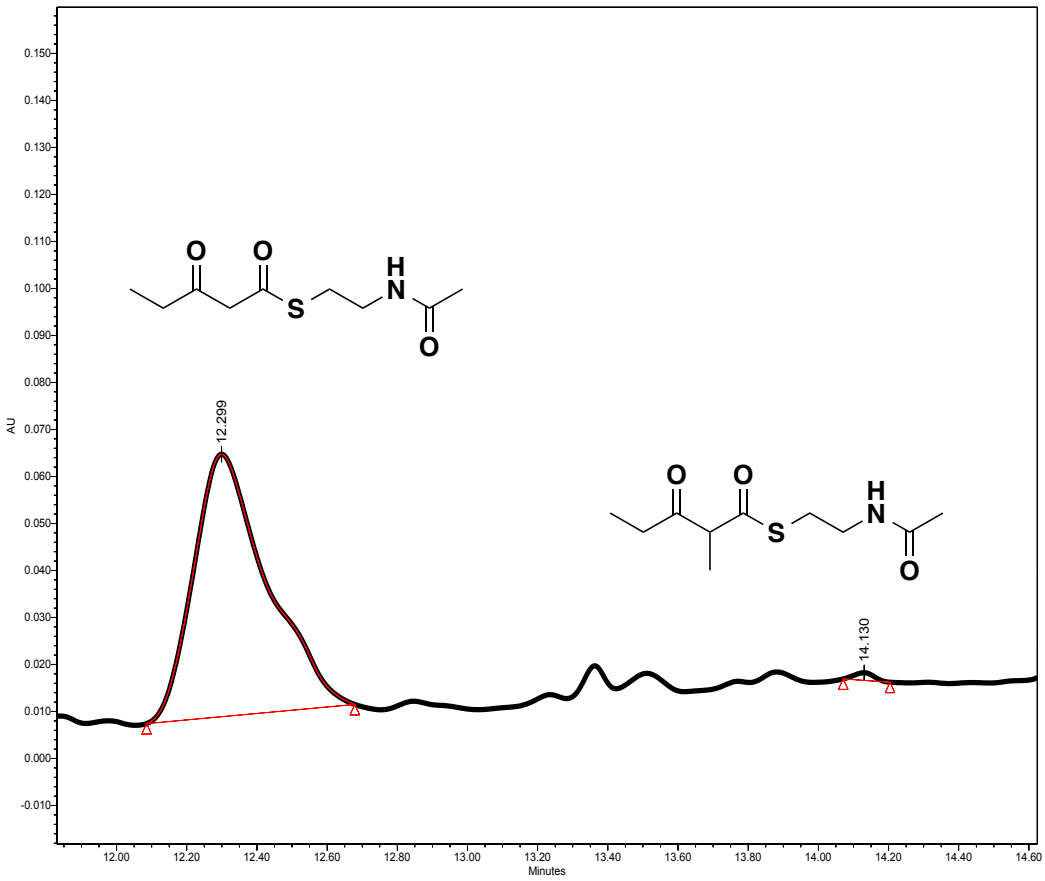


Figure 2-25: GphMT5 reaction HPLC trace 72 h with Pfs.

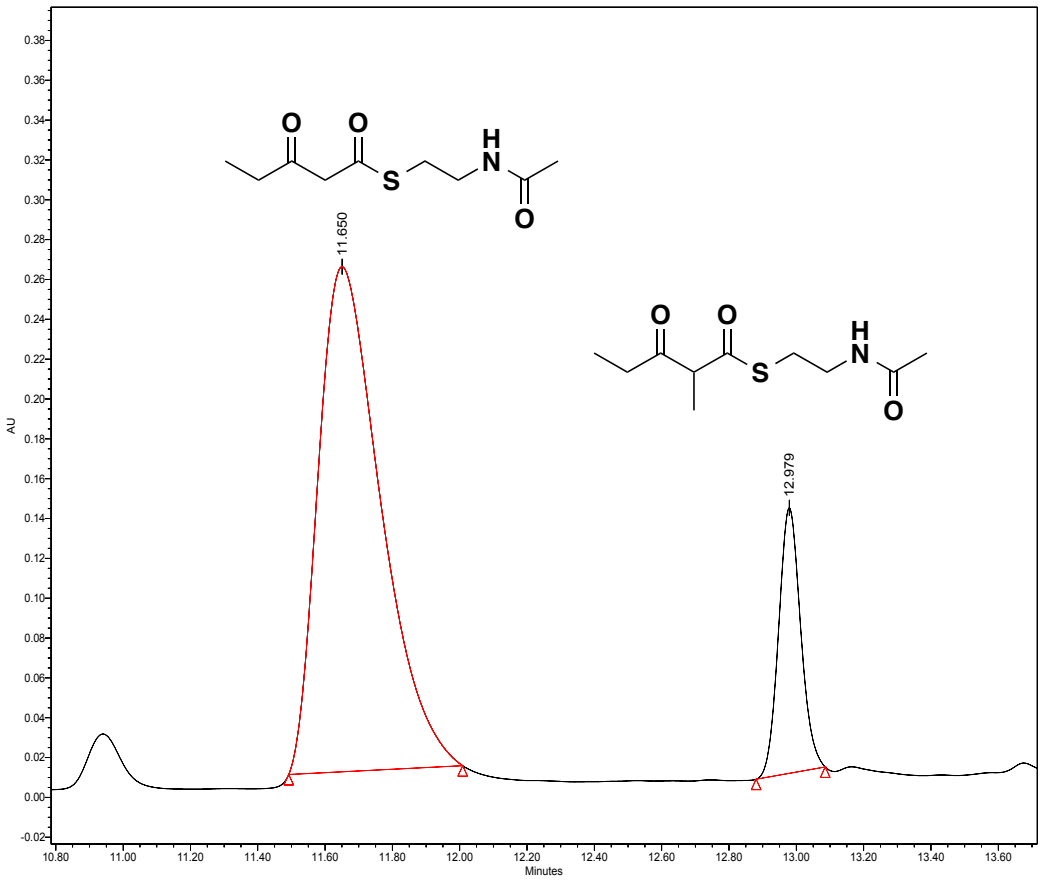


Figure 2-26: GphMT6 reaction HPLC trace 72 h without Pfs.

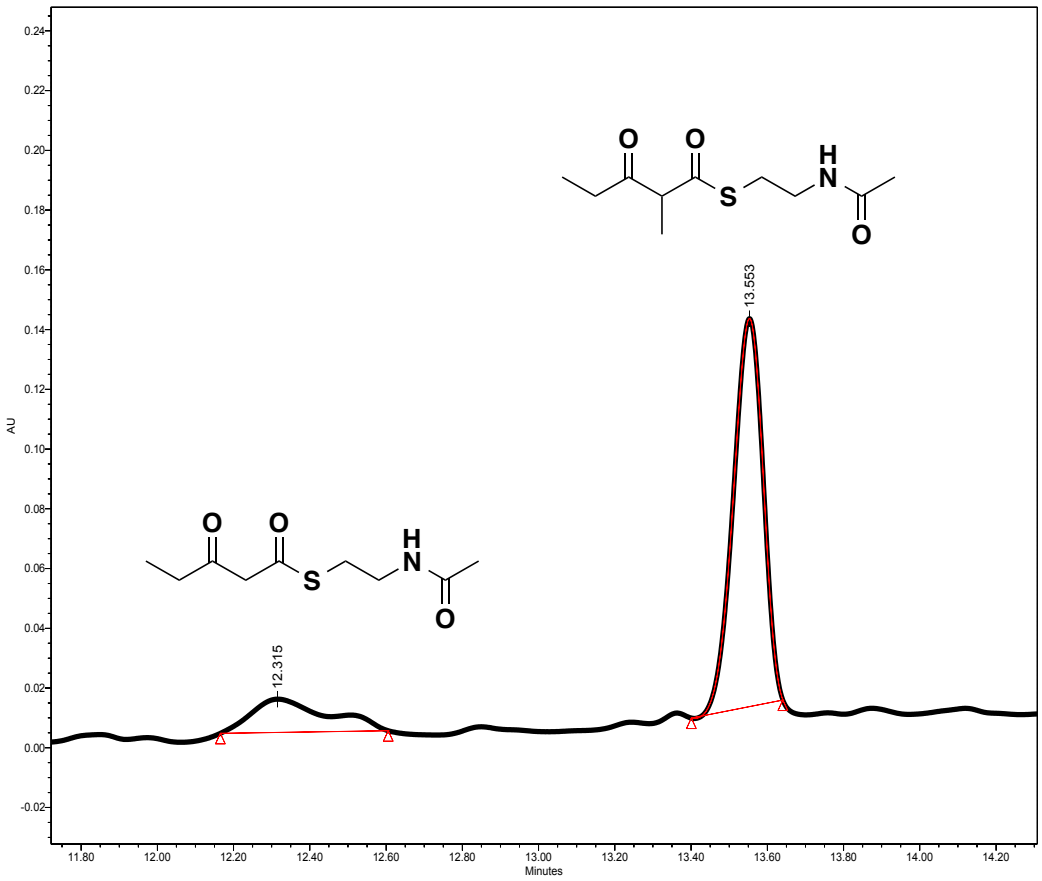


Figure 2-27: GphMT6 reaction HPLC trace 72 h with Pfs.

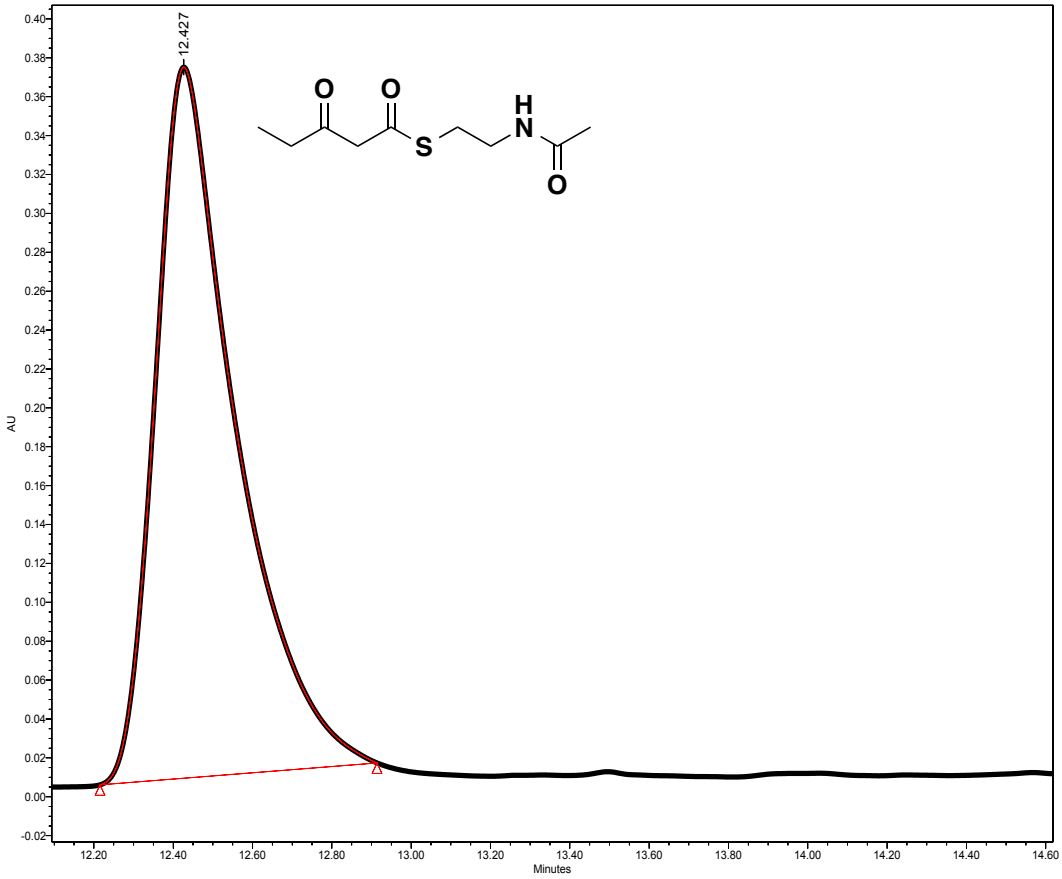


Figure 2-28: No Enzyme negative control HPLC trace 72 h without Pfs.

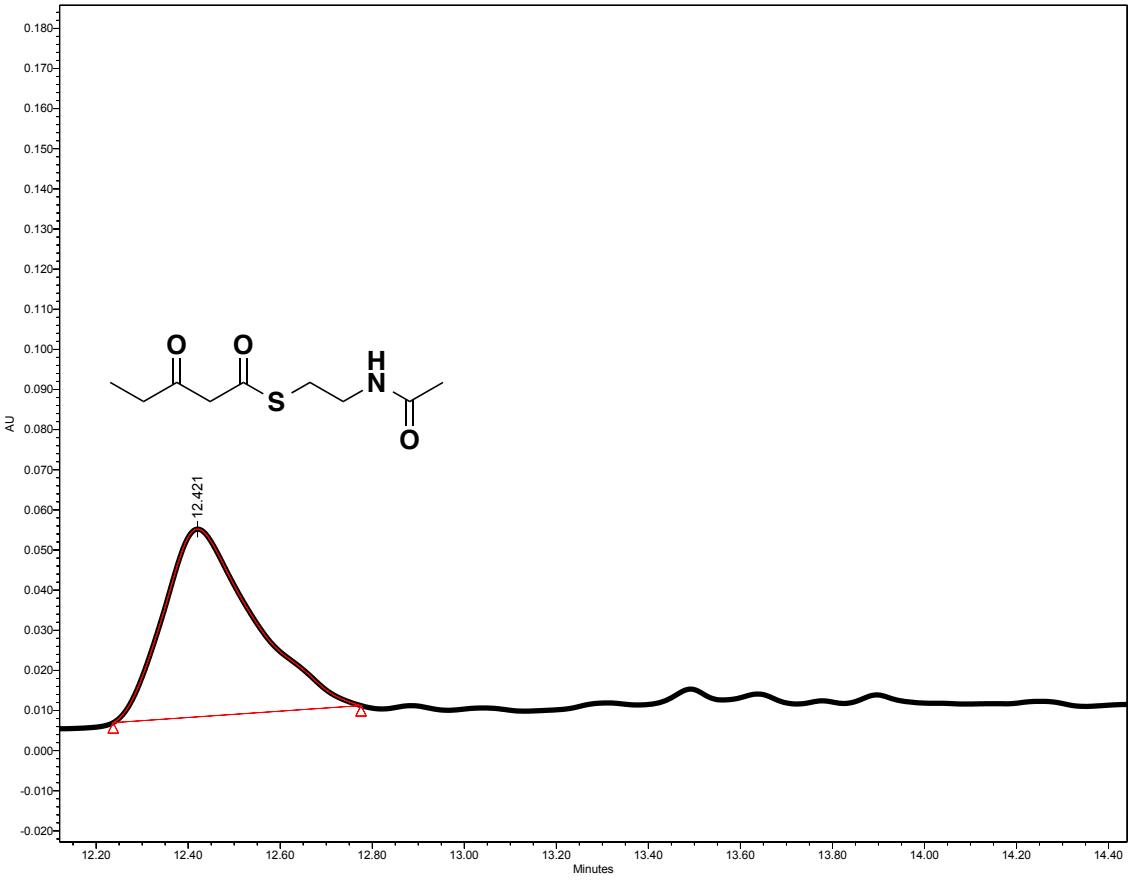
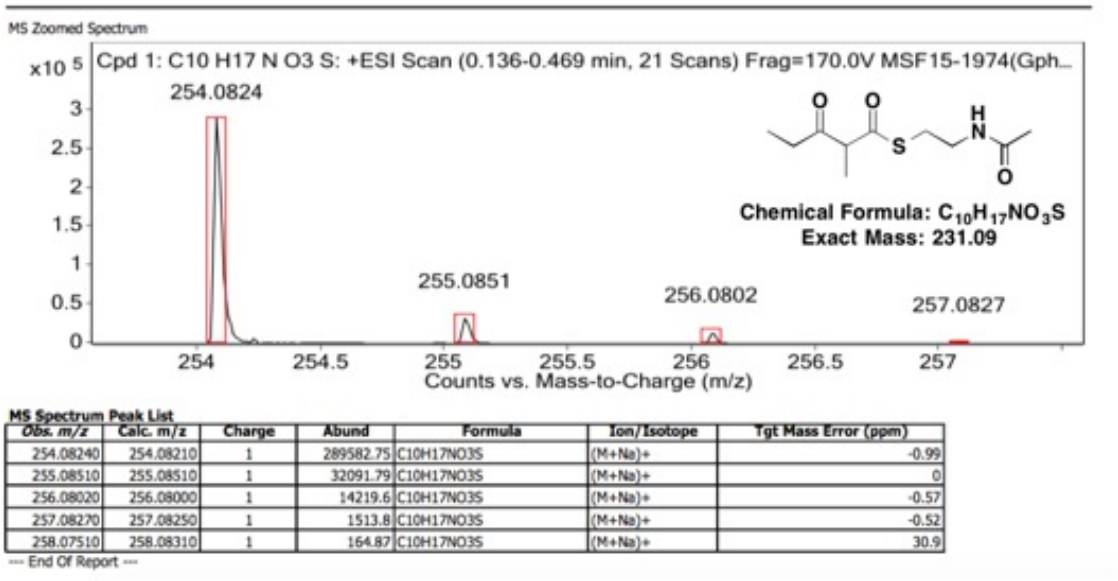


Figure 2-29: No Enzyme negative control HPLC trace 72 h without Pfs.

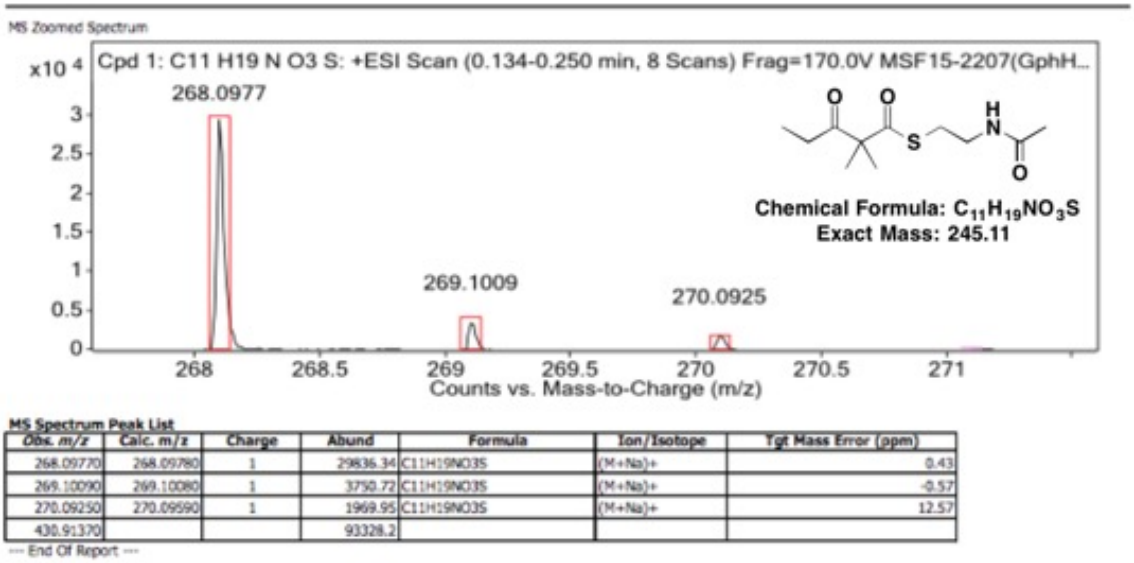


MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
254.08240	254.08210	1	289582.75	C10H17NO3S	(M+Na)+	-0.99
255.08510	255.08510	1	32091.79	C10H17NO3S	(M+Na)+	0
256.08020	256.08000	1	14219.6	C10H17NO3S	(M+Na)+	-0.57
257.08270	257.08250	1	1513.8	C10H17NO3S	(M+Na)+	-0.52
258.07510	258.08310	1	164.87	C10H17NO3S	(M+Na)+	30.9

--- End Of Report ---

Figure 2-30: HRMS of HPLC purified 2 generated from 1 via GphMT1.

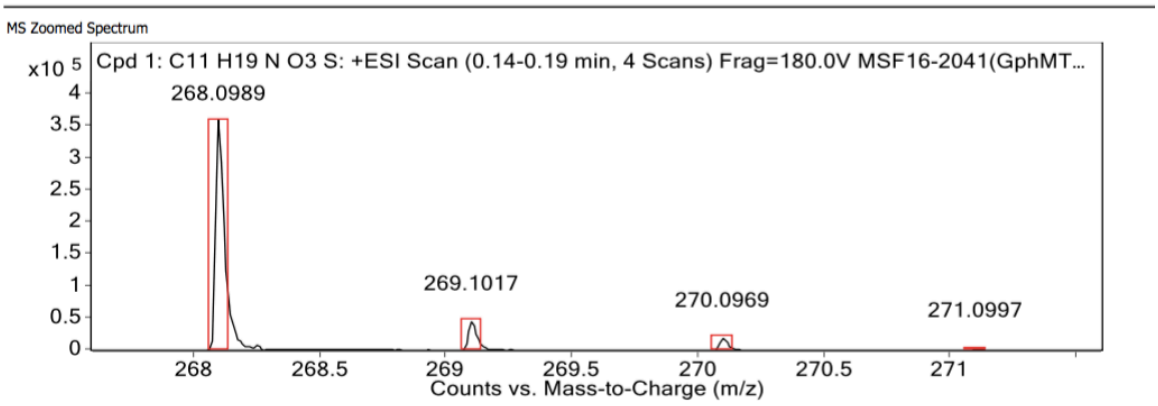


MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
268.09770	268.09780	1	29836.34	C11H19NO3S	(M+Na)+	0.43
269.10090	269.10080	1	3750.72	C11H19NO3S	(M+Na)+	-0.57
270.09250	270.09590	1	1969.95	C11H19NO3S	(M+Na)+	12.57
430.91370			93328.2			

--- End Of Report ---

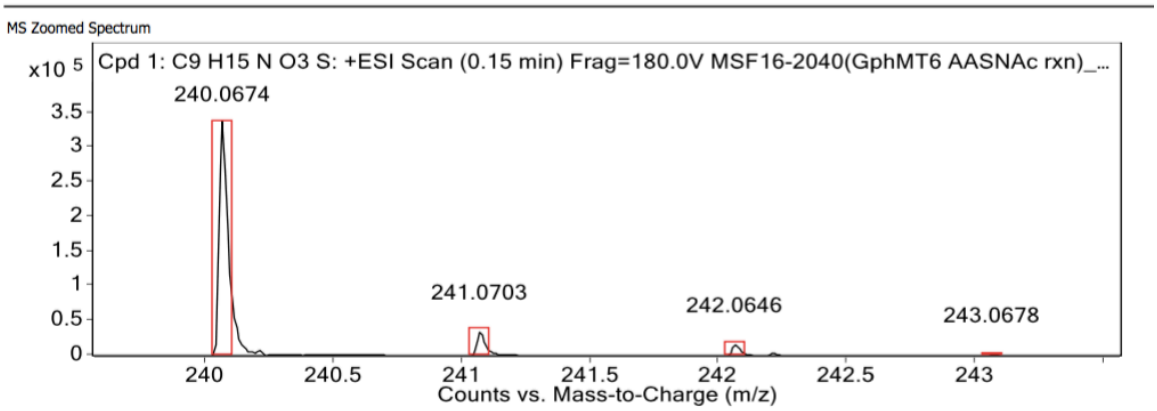
Figure 2-31: HRMS of HPLC purified 3 generated from 1 via GphMT5.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
268.09890	268.09780	1	360081.49	C11H19NO3S	(M+Na)+	-4.24
269.10170	269.10080	1	45052.04	C11H19NO3S	(M+Na)+	-3.53
270.09690	270.09590	1	18776.23	C11H19NO3S	(M+Na)+	-3.78
271.09970	271.09830	1	3095.82	C11H19NO3S	(M+Na)+	-5.28
272.09880	272.09900	1	471.19	C11H19NO3S	(M+Na)+	0.59

Figure 2-32: HRMS of HPLC purified 5 generated from 4 via GphMT6.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
240.06740	240.06650	1	337391.31	C9H15NO3S	(M+Na)+	-3.66
241.07030	241.06940	1	35095.04	C9H15NO3S	(M+Na)+	-3.77
242.06460	242.06420	1	16616.93	C9H15NO3S	(M+Na)+	-1.58
243.06780	243.06680	1	2126.64	C9H15NO3S	(M+Na)+	-4.29
244.06610	244.06720	1	494.45	C9H15NO3S	(M+Na)+	4.63

Figure 2-33: HRMS of HPLC purified 7 generated from 6 via GphMT6.

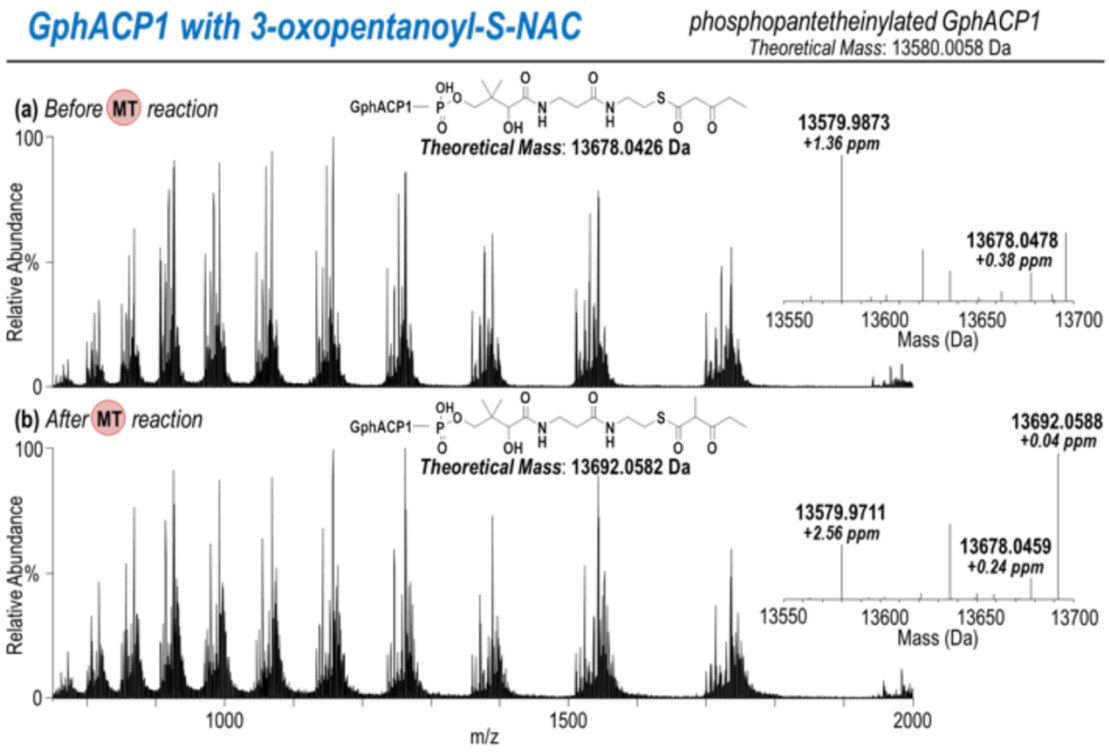


Figure 2-34: ESI-MS acyl- phosphopantetheinylated GphACP1.

ESI-MS of phosphopantetheinylated GphACP1 thioesterified with 3-oxopentanoyl-*S*-NAC (A) before and (B) after MT reaction. Insets give expanded views of the deconvoluted spectra (mass range 13550-13700 Da) labeled with the monoisotopic masses and mass accuracies relative to the theoretical mass of the modified proteins.

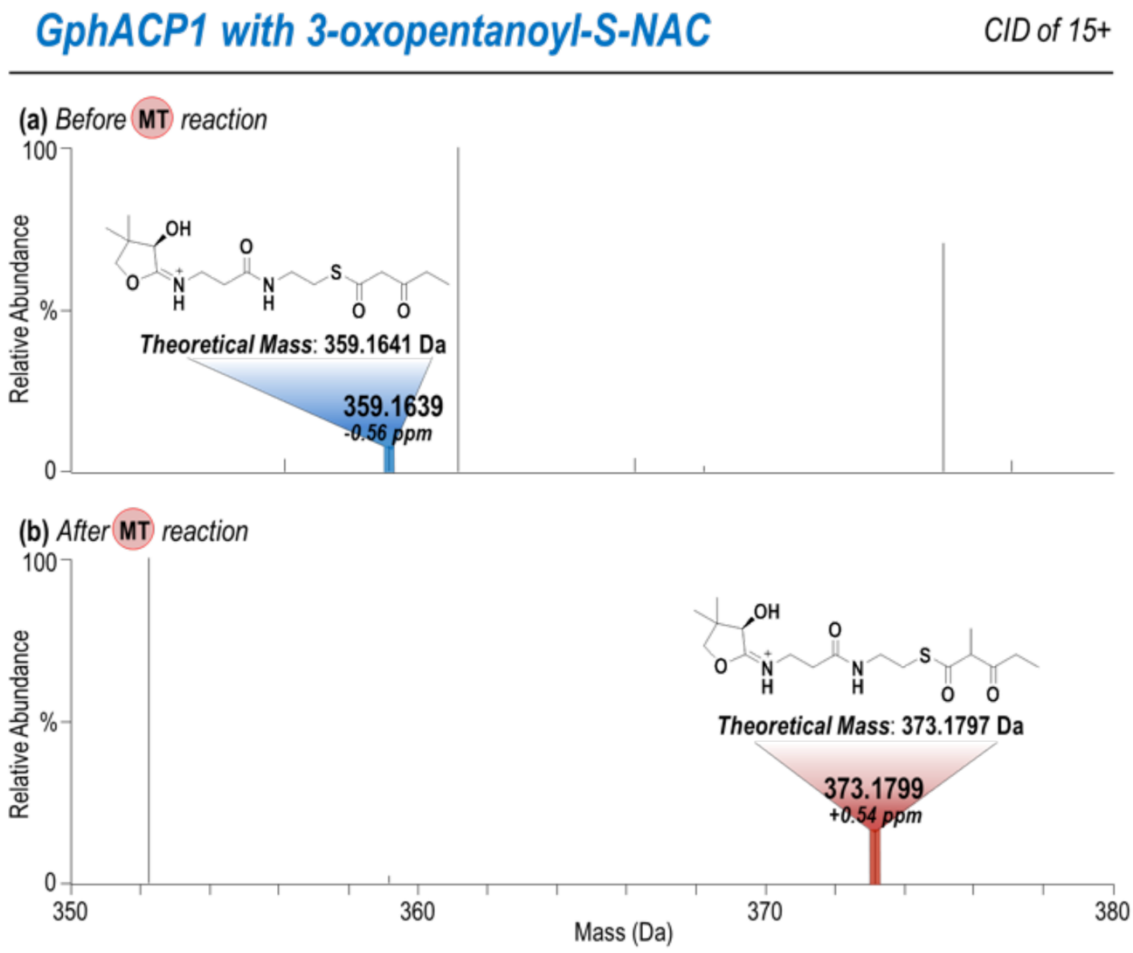


Figure 2-35: GphACP1 MS/MS before and after MT reaction.

Deconvoluted MS/MS spectrum (mass range 350-380 Da) resulting from CID of the 15+ charge state of phosphopantetheinylated GphACP1 with 3-oxopentanoyl-*S*-NAC (A) before and (B) after MT reaction. Highlighted peaks correspond to the thioesterified phosphopantetheine moiety cleaved from the protein upon collisional activation labeled with the monoisotopic mass and mass accuracy relative to the theoretical mass.

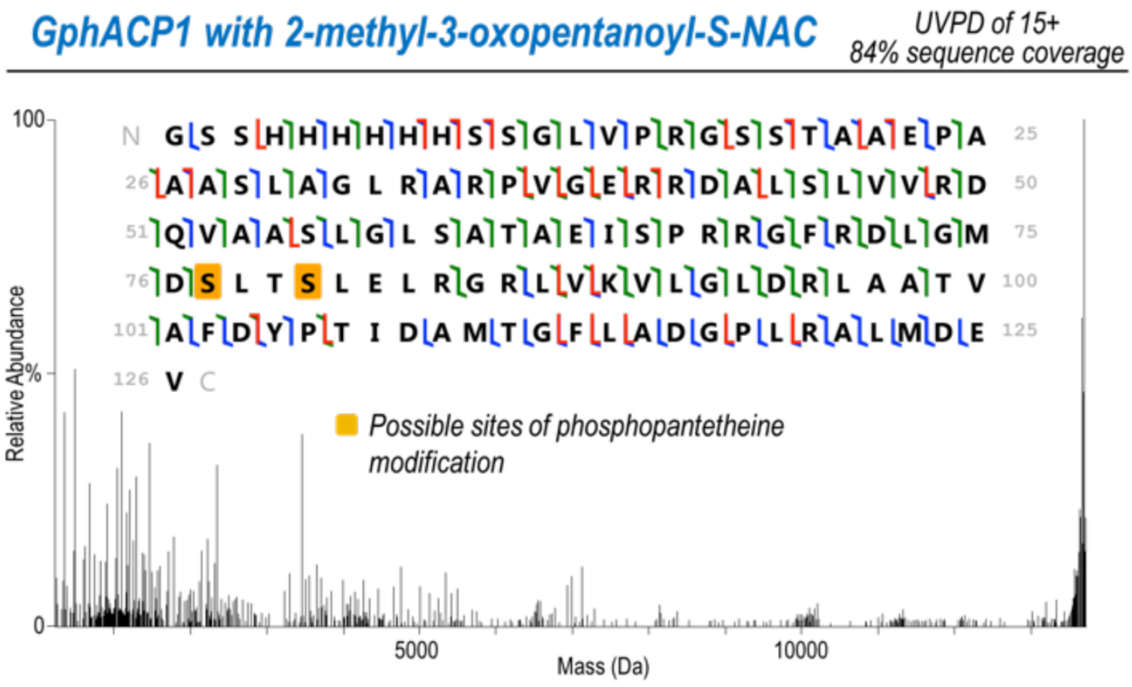


Figure 2-36: GphACP1 with 2-methyl-3-oxopentanoyl-*S*-NAC MS/MS.

Deconvoluted MS/MS spectrum resulting from UVPD of the 15+ charge state of phosphopantetheinylated GphACP1 with 2-methyl-3-oxopentanoyl-*S*-NAC. Two possible modification sites were identified (Ser77 or Ser80) based on mass shifts in the observed fragment ions.

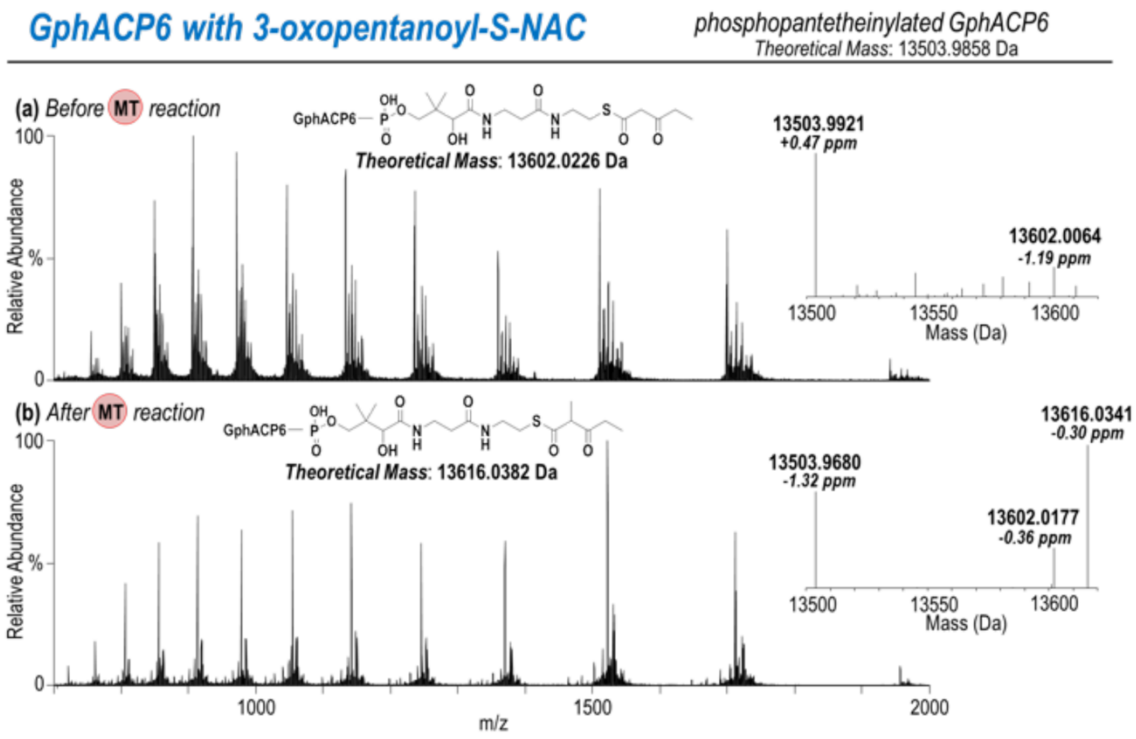


Figure 2-37: ESI-MS of phosphopantetheinylated GphACP6.

ESI-MS of phosphopantetheinylated GphACP6 thioesterified with 3-oxopentanoyl-*S*-NAC (A) before and (B) after MT reaction. Insets give expanded views of the deconvoluted spectra (mass range 13500-13620 Da) labeled with the monoisotopic masses and mass accuracies relative to the theoretical mass of the modified proteins.

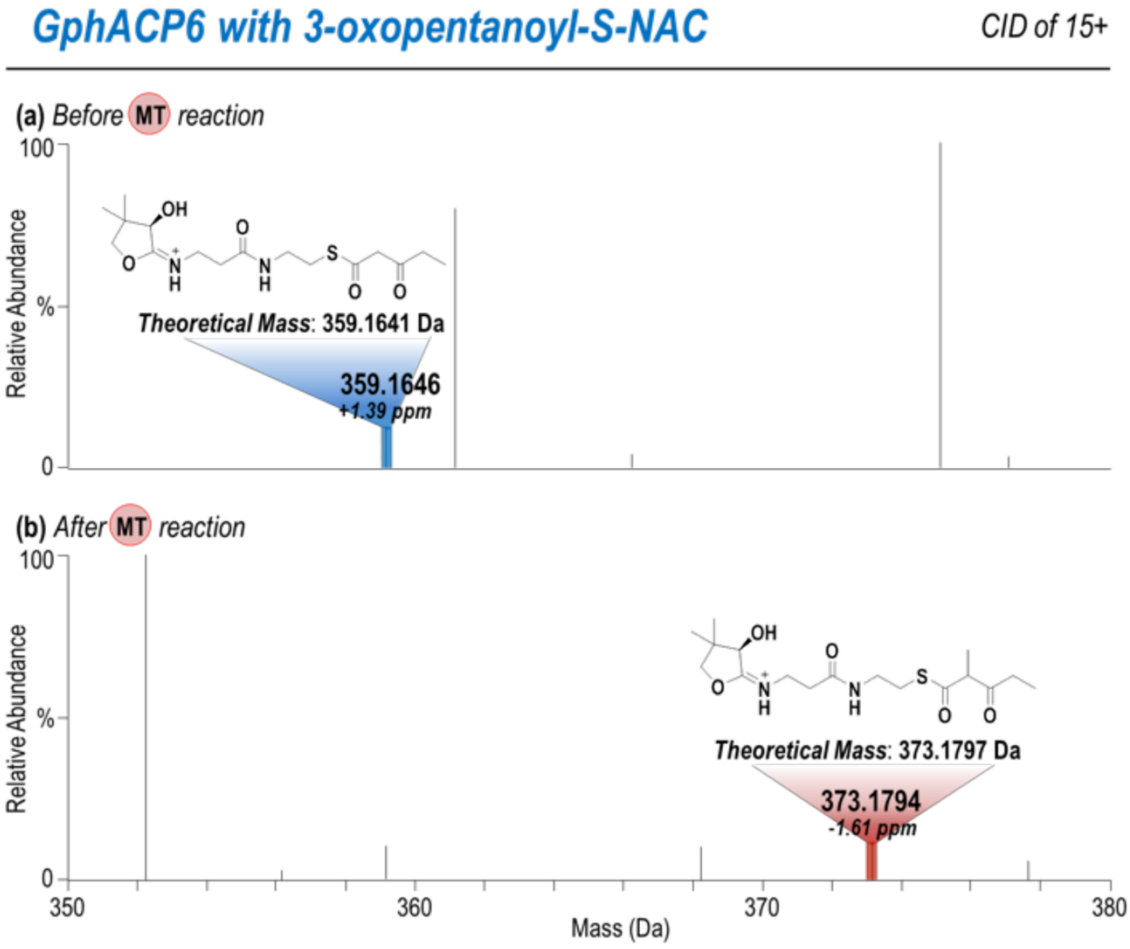


Figure 2-38**:** GphACP6 MS/MS before and after MT reaction.

Deconvoluted MS/MS spectrum (mass range 350-380 Da) resulting from CID of the 15+ charge state of phosphopantetheinylated GphACP6 with 3-oxopentanoyl-*S*-NAC (A) before and (B) after MT reaction. Highlighted peaks correspond to the thioesterified phosphopantetheine moiety cleaved from the protein upon collisional activation labeled with the monoisotopic mass and mass accuracy relative to the theoretical mass.

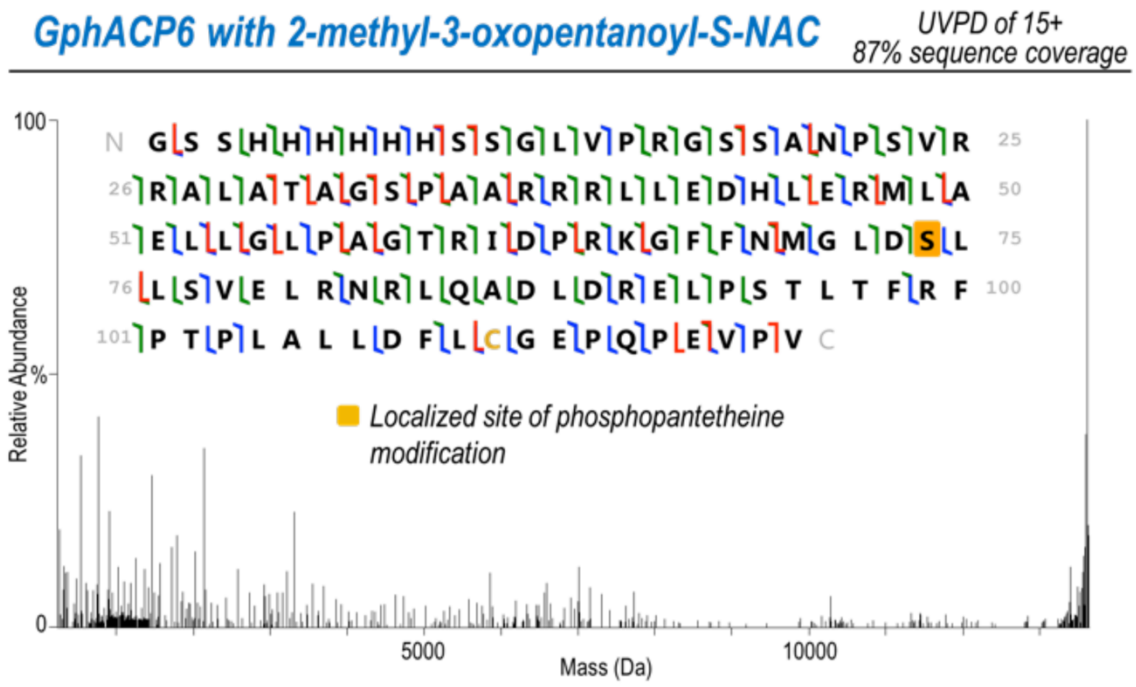


Figure 2-39: GphACP6 with 2-methyl-3-oxopentanoyl-*S*-NAC MS/MS.

Deconvoluted MS/MS spectrum resulting from UVPD of the 15+ charge state of phosphopantetheinylated GphACP6 with 2-methyl-3-oxopentanoyl-*S*-NAC. The modification was localized to Se74 based on mass shifts in the observed fragment ions.

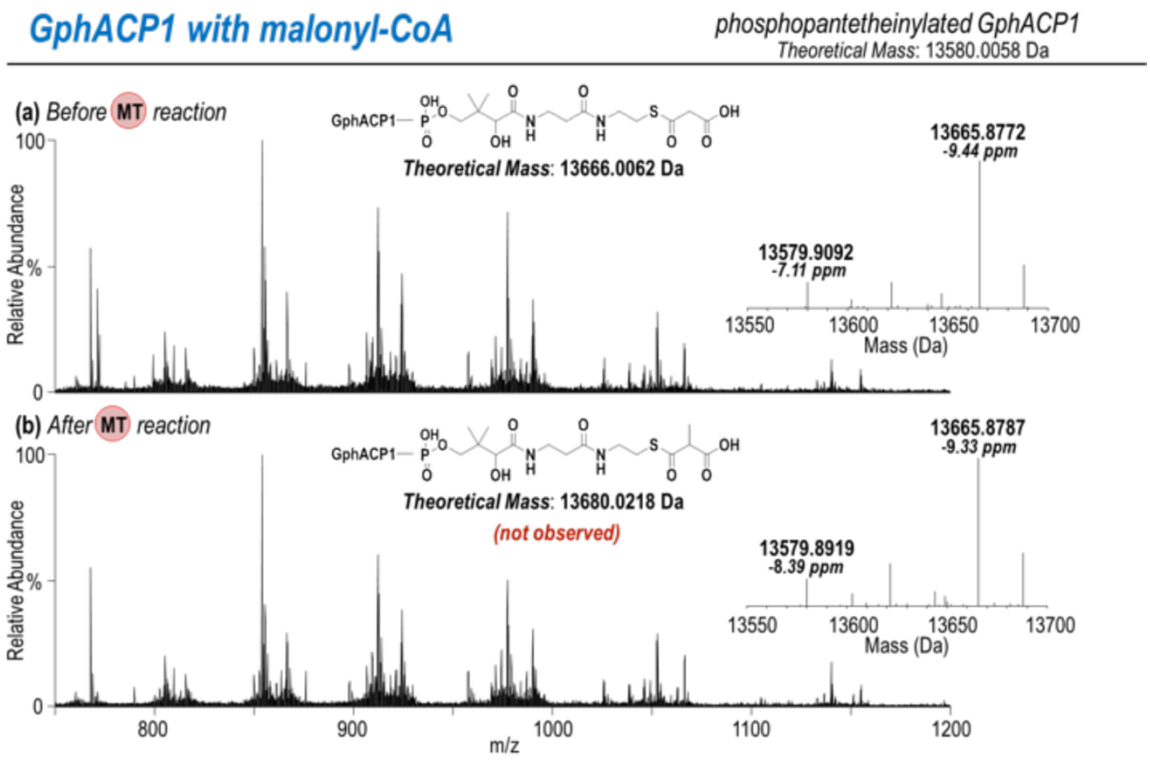


Figure 2-40**:** ESI-MS of malonyl-phosphopantetheinylated GphACP1.

ESI-MS of phosphopantetheinylated GphACP1 thioesterified with malonyl-CoA (A) before and (B) after MT reaction. The methylated form was not observed after the reaction. Insets give expanded views of the deconvoluted spectra (mass range 13550-13700 Da) labeled with the monoisotopic masses and mass accuracies relative to the theoretical mass of the modified proteins.

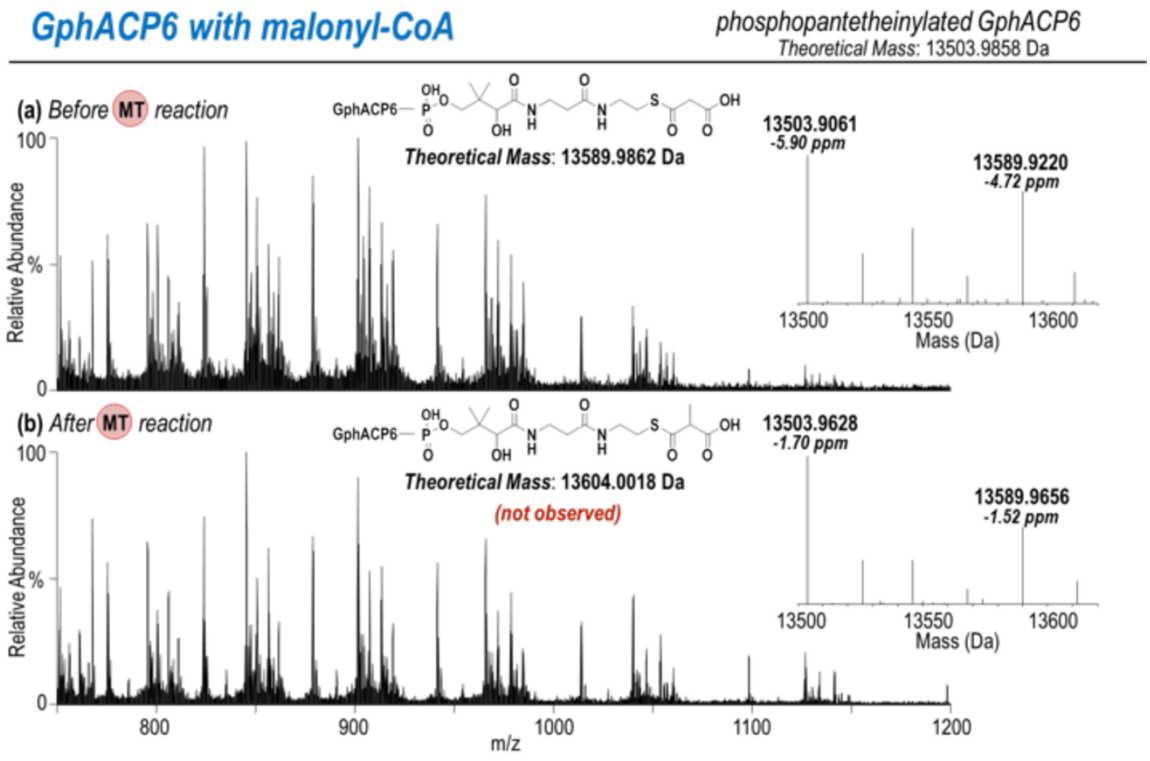


Figure 2-41**:** ESI-MS of malonyl-phosphopantetheinylated GphACP6.

ESI-MS of phosphopantetheinylated GphACP6 thioesterified with malonyl-CoA (A) before and (B) after MT reaction. The methylated form was not observed after the reaction. Insets give expanded views of the deconvoluted spectra (mass range 13500-13620 Da) labeled with the monoisotopic masses and mass accuracies relative to the theoretical mass of the modified proteins.

Amplicon	Forward Primer (5' -> 3')	Reverse Primer (5' → 3')
GphMT1	GCGGCCTGGTGCCGCGCGGCTCTAGCGCCGCGCTGCTCGACTC	GTGGTGGTGGTGGTGGTGATGTTAAACCGGACGGGCGCGG
GphMT2	GCGGCCTGGTGCCGCGCGGCTCTAGCGACGTCTCGCTGTTCGACG	GTGGTGGTGGTGGTGGTGATGTTACGCGCGGATCCGG
GphMT3	GCGGCCTGGTGCCGCGCGGCTCTAGCACGGCTGTGCAGGAGCCAC	GTGGTGGTGGTGGTGGTGATGTTATTGCGGGAGGCCACCCA
GphMT4	GCGGCCTGGTGCCGCGCGGCTCTAGCGCCGAGGGGCTCCAAGG	GTGGTGGTGGTGGTGGTGATGTTACGATCCGCGCGACCAG
GphMT5	GCGGCCTGGTGCCGCGCGGCTCTAGCGCCCCCGGCACTTTCG	GTGGTGGTGGTGGTGGTGATGTTACTCGAGCGTGGCGAGCC
GphHmod	GCGGCCTGGTGCCGCGCGGCTCTAGCATGAGCAACGGGCAGTCGG	GTGGTGGTGGTGGTGGTGATGTTATCACGTGCCCAGGAGCG
GphMT6	GCGGCCTGGTGCCGCGCGGCTCTAGCGCCGTTGGCGCCGTC	GTGGTGGTGGTGGTGGTGATGTTACGCGGCGGGACTCCC
GphACP1	GCGGCCTGGTGCCGCGCGGCTCTAGCACGGCGGCTGAGCCC	GTGGTGGTGGTGGTGGTGATGTTACACCTCGTCCATCAGCGC
GphACP6	GCGGCCTGGTGCCGCGCGGCTCTAGCGCCAACCCGTCCGTGC	GTGGTGGTGGTGGTGGTGATGTTACACCGGGACCTCGGG
Pfs	GCGGCCTGGTGCCGCGCGGCTCTAGCATGAAAATCGGCATCATTGGTG	GTGGTGGTGGTGGTGGTGATGTTAGCCATGTGCAAGTTTCTGCAC

Table 2-1**:** Primer sequences used for construction of all MT plasmids.

Chapter 3: Methyltransferases Excised From trans-AT Polyketide Synthases Operate on N-acetylcysteamine-bound Substrates

Adapted from Stevens, et al. (2016) *J. Antibiot (Tokyo)*. **69**, 567-570.

ABSTRACT

Trans-acyltransferase polyketide synthases are modular assembly lines employed by microbes to generate diverse, bioactive products. The methyl groups that branch from the polyketide scaffolds of these natural products are installed by synthase-embedded methyltransferases. While these methyltransferases greatly broaden the chemical space available to the polyketide natural products, they remain one of the least investigated processing enzymes of the modular polyketide synthases. Herein we examine the activity and selectivity of methyltransferases excised from the bacillaene, difficidin, and mupirocin synthases towards β-ketoacyl-*S*-*N*-acetylcysteamine analogs of their natural substrates. The results provide the first steps in the characterization of this class of methyltransferase as well as their development as biocatalysts.

INTRODUCTION

Trans-acyltransferase polyketide synthases (*trans*-AT PKSs) construct biologically active polyketides, such as the antibiotic mupirocin, in an assembly line process catalyzed by collections of enzymatic domains termed modules. Each module of a *trans*-AT possesses an acyl carrier protein (ACP) and ketosynthase (KS) that, along with a separately-encoded acyltransferase (AT) domain, constitute the minimal trio of

domains required for the extension of a covalently-bound intermediate (Piel, 2010; Helfrich et al., 2016). Further modifications at the α- and β-carbons of the polyketide are introduced by processing enzymes such as the methyltransferase (MT), ketoreductase (KR), dehydratase (DH), and enoylreductase (ER), depending on their inclusion within a module. This optional incorporation can be primarily credited for the vast chemical and functional diversity observed within polyketides.

Unlike the embedded AT domains of *cis*-AT PKS pathways that deliver malonyl-CoA or methylmalonyl-CoA extender units to a single module, the discrete ATs of *trans*-AT PKSs deliver extender units to several modules. Since these ATs are primarily selective for malonyl groups, *trans*-AT PKSs employ embedded *S*-adenosyl methionine (SAM)-dependent MT domains to install methyl groups into polyketide backbones (Piel, 2010; Helfrich et al., 2016; Keatinge-Clay, 2012). Although SAM-dependent MTs constitute a well-studied superfamily of enzymes and are relatively common within *trans*-AT PKSs, a dearth of information exists for *trans*-AT MTs and PKS-embedded MTs in general. Only a few studies of MTs from select *cis*-AT PKSs and fungal highly-reducing PKSs (HR-PKSs) have been reported (Ansari et al., 2008; Lipscombe et al., 2012, Miller et al., 2002; Poust et al., 2015). Analysis of a fungal HR-PKS responsible for the production of lovastatin revealed the kinetic preference of an embedded MT towards its native acyl-*S*-*N*-acetylcysteamine (SNAC) substrate analog (Miller et al., 2015; Jenner at al., 2013; Jenner at al., 2015).

To explore the selectivity of *trans*-AT MTs, we examined the methylation activity of excised MT domains from three well-known *trans*-AT pathways, responsible for the

production of the antimicrobial agents bacillaene, difficidin, and mupirocin (Figure 3-1), towards the acyl-SNAC substrates 3-oxobutanoyl-*S*-NAC (**1**), 3-oxopentanoyl-*S*-NAC (**2**), and 3-oxohexanoyl-*S*-NAC (**3**) (Moldenhaurer et al., 2010; Chen et al., 2006; Gurney et al., 2011). (Table 3-1). This is the first account of the substrate specificity and activity of excised MT domains from *trans*-AT PKSs.

RESULTS AND DISCUSSION

Expression of MT Domains

Of the 7 domains cloned, all produced soluble protein with the exception of MupMT3. In addition to several previously identified sequence motifs that distinguish between *cis*- and *trans*-AT MT domains, a *trans*-AT MT sequence alignment revealed an additional notable difference between these two classes of MT domain regarding their position of within the module (Young et al., 2013). In contrast to their *cis*-AT counterparts, *trans*-AT MTs are not embedded within the KR structural subdomain but are instead usually situated immediately after the KR domain and immediately before the ACP domain (Nguyen et al, 2008).

Catalytic Activity of Excised MT Domains

In vitro assays of the excised MTs BaeMT8, DifMT1, and MupMT1 showed each to be capable of catalyzing the methylation of substrates **1, 2,** and **3** to afford **4, 5,** and **6** with 10-100% conversion observed after 16 h at 1-5 mM substrate concentrations (Figure 3-2). These results show that, at least when separated from their assembly lines, MTs are not highly substrate-specific. Other excised PKS enzymes [KRs, DHs, ERs and

thioesterases (TEs)] also demonstrate broad substrate specificities (Young et al., 2013; Siskos et al., 2005). The low reaction rates of BaeMT8, DifMT1, and MupMT1 are suggestive that local substrate concentrations within the intact assembly line are required for optimal catalytic rates (Table 1). While the conversions of **1** to **4** catalyzed by BaeMT8 and DifMT1 as well as **2** to **5** catalyzed by BaeMT8 were observed overnight, the percent conversions within the linear region (1-3h) were too low to accurately report kinetic parameters using HPLC analysis. The observed reaction rates for MTs are lower than those of other excised processing domains (e.g., KRs, DHs, or TEs) towards *N*-acetylcysteamine substrates (Li et al., 2014; Wang et al., 2008; Li et al., 2015; Zheeg et al., 2012). Although DifMT1 and BaeMT8 demonstrated substrate affinities for **2** and **3** respectively (Table 3-1), a significant kinetic preference for a specific substrate was not observed from any of the investigated MTs.

The N-terminal MTs DifMT6, DifMT13, and BaeMT12 did not catalyze a detectable amount of methylation of **1**, **2**, or **3**. A partial misfold of the protein may have attenuated activity, this result suggests that, unlike the completely embedded MTs, N-terminal MTs require additional components for activity, such as inclusion in their native polypeptide or interaction with the upstream PKS polypeptide.

MATERIALS AND METHODS

Materials

S-adenosyl-methionine (Ark Pharm, Inc.) was dissolved in 300 mM sodium phosphate buffer (pH 7.8) to a final concentration of 150 mM.

Cloning and Expression of MT Domains

The DNA encoding DifMT1, DifMT6, DifMT13, BaeMT8, BaeMT12, MupMT1, and MupMT3 was amplified from the difficidin and bacillaene gene clusters of *Bacillus amyloliquefaciens* FZB42 and the mupirocin gene cluster of *Pseudomonas fluorescens* 13525 and inserted into pGAY28b, a ligation-independent cloning vector constructed from pET28b (Table 3-2) (Gay et al., 2015. *E. coli* BL21(DE3) transformed with the expression plasmid was inoculated into LB media containing 50 mg/L kanamycin at 37 °C, grown to OD_{600} = 0.5, and induced with 0.5 mM IPTG. After 18 h at 15 °C, cells were collected by centrifugation and resuspended in lysis buffer (0.5 M NaCl, 10% (v/v) glycerol, 0.1 M HEPES, pH 7.5). Following sonication, cell debris was removed by centrifugation (30,000 x g, 30 min). The supernatant was poured over a column of Nickel-NTA resin (Thermoscientific), which was then washed with 40 mL lysis buffer containing 15 mM imidazole and eluted with 5 mL lysis buffer containing 150 mM imidazole. The eluted protein was concentrated to ~10 mg/mL in the equilibration buffer and stored at -80 ˚C until needed. The *S*-adenosyl-homocysteine (SAH) nucleosidase Pfs was amplified from *E. coli* BL21(DE3) genomic DNA and cloned and purified as above (Supplementary Table 1). Reactions were supplemented with Pfs due to the potent inhibition of MTs by SAH (Liscombe et al., 2012).

Synthesis of Acyl-SNAC Substrates

Compounds **1**-**3**, 2-methyl-3-oxobutanoyl-*S*-NAC (**4**), and 3-oxohexanoyl-*S*-NAC (**6**) were prepared according to reported protocols, as was 2-methyl-3-oxopentanoyl-*S*-NAC (**5**) (Piasecki et al., 2011; Bailey at el., 2016).

Kinetic Analysis of Methylation Activity

To determine MT activity reactions with substrates **1**-**3** (10 mM) contained 150 mM Tris-HCl, 100 mM NaCl, 20 mM SAM, 10% (v/v) glycerol, 20 uM Pfs, and 10 uM MT in a total volume of 200 uL. Reactions were run in parallel at 25 °C for ~16h. Standards of **4, 5,** and **6** at concentrations from 250 uM-50 mM were added directly to injection solvent and HPLC peak areas were collected to provide standard curves for observable product formation.

To determine the linear region of MT kinetic data reactions with substrates **1**-**3** (10 mM) contained 150 mM Tris-HCl, 100 mM NaCl, 20 mM SAM, 10% (v/v) glycerol, 20 ¦ M Pfs, and 10 uM MT in a total volume of 200 uL and were run in parallel at 25 °C and quenched at Oh, 1h, 2h, 3h, 4h, and 5h time points.

Reactions with substrates **1**-**3** (100 uM, 500 uM, 1mM, 2.5 mM, 5 mM, 10 mM, or 50 mM) contained 150 mM Tris-HCl, 100 mM NaCl, 20 mM SAM, 10% (v/v) glycerol, 20 uM Pfs, and 10 uM MT in a total volume of 200 uL. Reactions were run in parallel at 25 °C over a period of 1-3 hours (MupMT1+**1**, 1h; MupMT1+**2**, 2h; MupMT1+**3**, 3h; BaeMT8+**3**, 2h; DifMT1+**2**, 2h; DifMT1+**3**, 2h). After completion, reactions were extracted with 1 mL of ethyl acetate. The organic layer was dried *in vacuo* and resuspended in methanol for HPLC analysis. Standard curves were generated by HPLC analysis of 60 uM, 300 uM, 600 uM, 1 mM, 3 mM, and 6 mM injections of **4** and 200 uM, 1 mM, 2 mM, 5 mM, 10 mM, and 20 mM injections of **5** and **6.** Reactions for each set of substrate concentrations were performed in triplicate.

Product formation was quantified by absorbance with reverse phase HPLC resolution of reaction mixtures stopped at a fixed time-point with 0.1 – 50 mM substrate as well as full progress curves at 10 mM substrate (Figure 3-2 – 3-13). Global fitting of both data sets using using Kintek Explorer allowed determination of k_{cat}/K_M values for each enzyme/substrate pair (Johnson and Simpson, 2009, Johnson et al., 2009).

HPLC and HRMS Analysis

HPLC analysis was performed with a tandem Waters 2707 autosampler and Waters 1525 binary HPLC pump connected to a Waters 2998 photodiode array detector using a Varian Microsorb-MV C_{18} column (250 x 4.6 mm, 5 μm particle size, 100 Å pore size) and mobile phases consisting of water with 0.1% TFA (solvent A) and acetonitrile with 0.1% TFA (solvent B) with a solvent gradient of 5%-100% B over 30 min at a flow rate of 1 mL/min. High-resolution mass spectrometry measurements were obtained by chemical ionization (ESI) with a VG analytical ZAB2-E instrument. Peak integrations were automated with a signal-to-noise cutoff of 5% peak area.

CONCLUSION

The presented work demonstrates that, similar to the other processing domains excised from PKSs, excised *trans*-AT PKS MTs are capable of operating on thioester-bound polyketide substrates *in vitro*. These data disclose the initial investigation of the heretofore unexplored MTs found throughout the *trans*-AT PKS landscape and provide insight into the catalytic activity of the enzyme when removed from its native module.

The ability of these MTs to install methyl branches on the carbon chains of small molecule substrates makes them particularly attractive as biocatalysts.

ACKNOWLEDGEMENTS

We would like to thank Dr. Kenneth Johnson for helpful discussion pertaining to the fitting of kinetic data. We would also like to thank the University of Texas at Austin mass spectrometry facility for their help in obtaining high-resolution masses for **1**, **2**, **3**, **4**, **5**, **6**. We thank the National Institutes of Health, National Institute of General Medical Sciences (GM106112) and the Welch Foundation (F-1712) for funding.

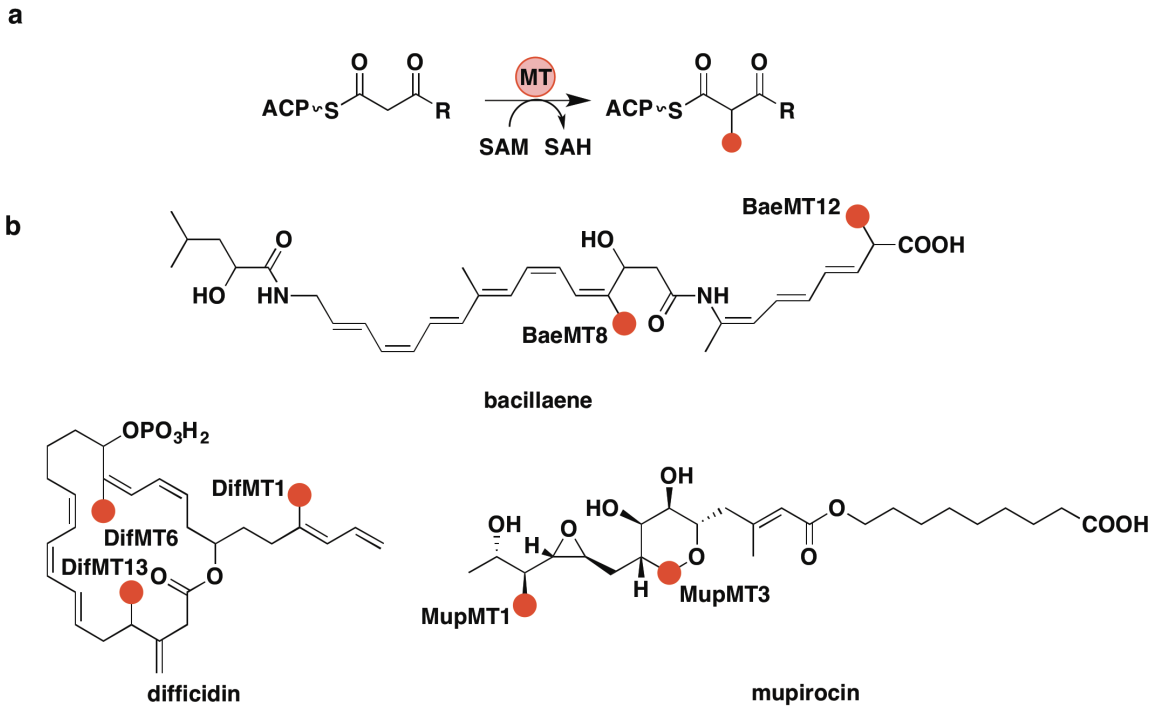


Figure 3-1: α-methylation catalyzed by *trans*-AT PKS MTs.

(A) Canonical alpha-methylation catalyzed by trans-AT PKS MTs.

(B) Methyl branches and cognate MT domains of bacillaene, difficidin, and mupirocin.

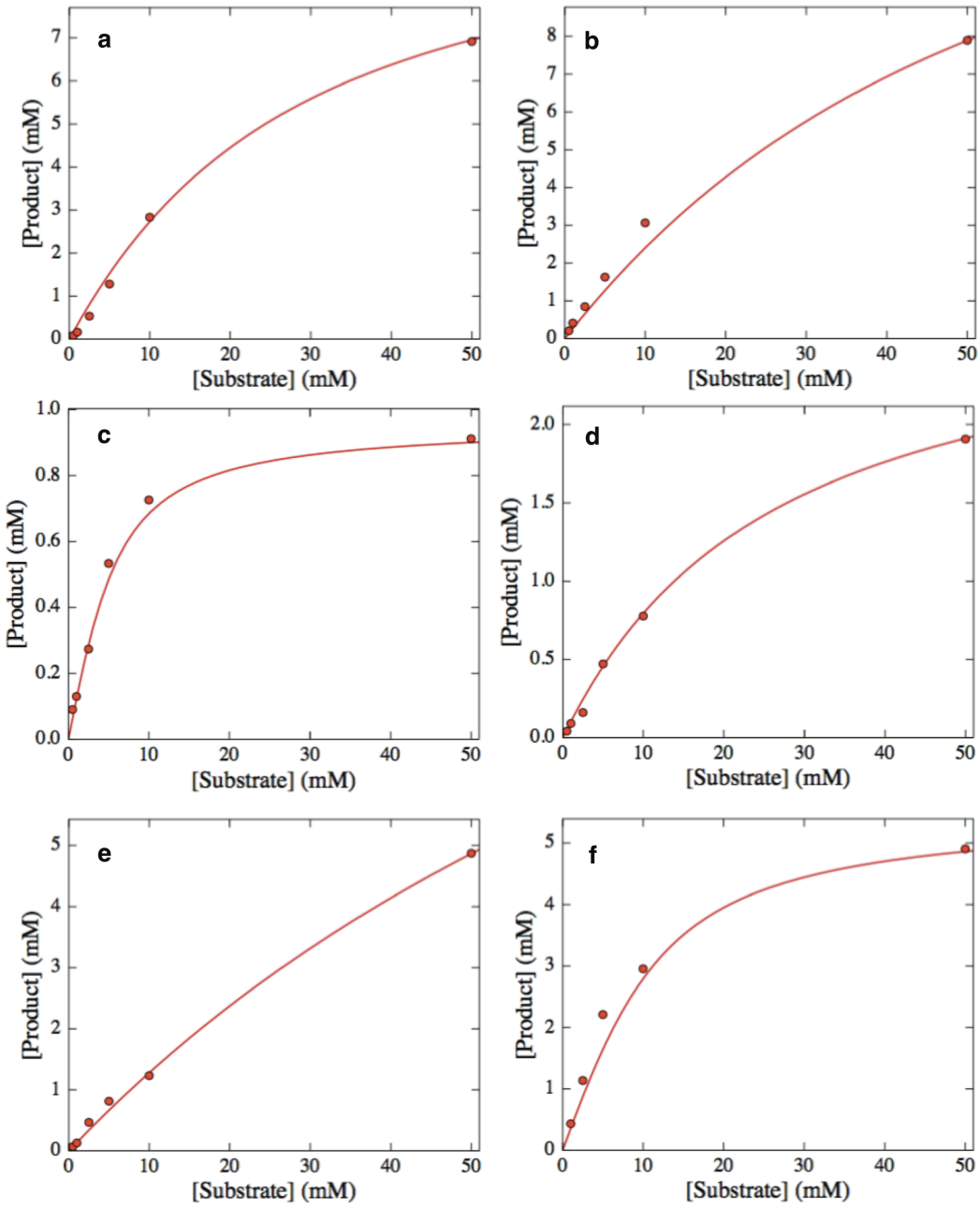


Figure 3-2: Methyltransferase saturation curves.

Figure 3-2: continued;

Saturation curves for a) MupMT1 and 1, b) MupMT1 and 2, c) MupMT1 and 3, d) BaeMT8 and 3, e) DifMT1 and 2, f) DifMT1 and 3. Experimental data from saturation curves globally fit to a fixed concentration time-course for enzyme-substrate pairs.

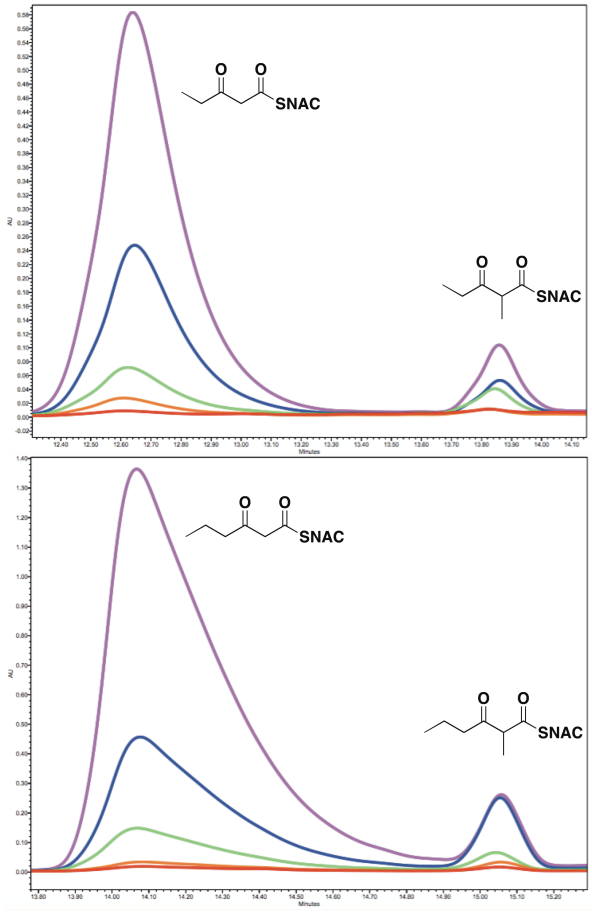


Figure 3-3: DifMT1 methylation activity of 3-oxopentoyl-S-NAC.

HPLC traces monitoring 235 nm for, Top set: DifMT1 catalyzed methylation of 3-oxopentanoyl-*S*-NAC (**2**) to afford 2-methyl-3-oxopentanoyl-*S*-NAC **(5)** at 500 uM (red) 1 mM(orange), 2.5 mM (green), 5 mM (blue), 10 mM (violet) at 2h and Bottom set: DifMT1 catalyzed methylation of 3-oxohexanoyl-*S*-NAC (**3**) to afford 2-methyl-3-oxohexanoyl-*S*-NAC (**6**) at 500 uM (red), □ 1 mM(orange), 2.5 mM (green), 5 mM (blue), 10 mM (violet) at 2h.

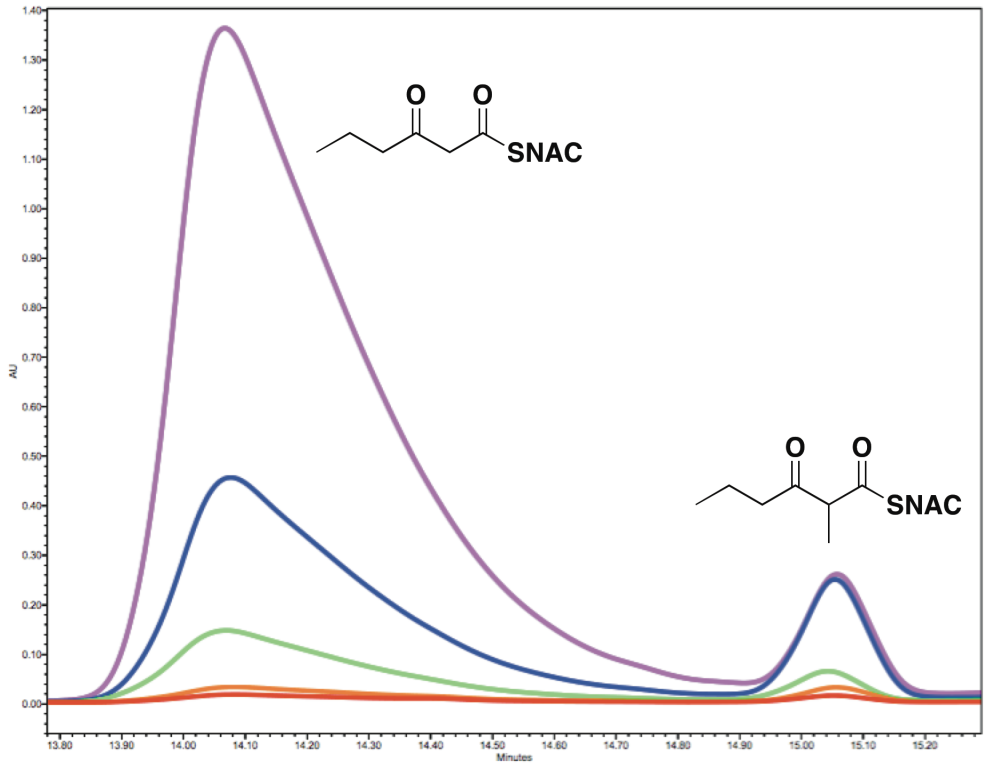


Figure 3-4: BaeMT8 methylation activity of 3-oxohexanoyl-*S*-NAC.

HPLC traces monitoring 235 nm for BaeMT8 catalyzed methylation of 3-oxohexanoyl-*S*-NAC (**3**) to afford 2-methyl-3-oxohexanoyl-*S*-NAC **(6)** at 500 uM (red), 1 mM (orange), 2.5 mM (green), 5 mM (blue) at 2h.

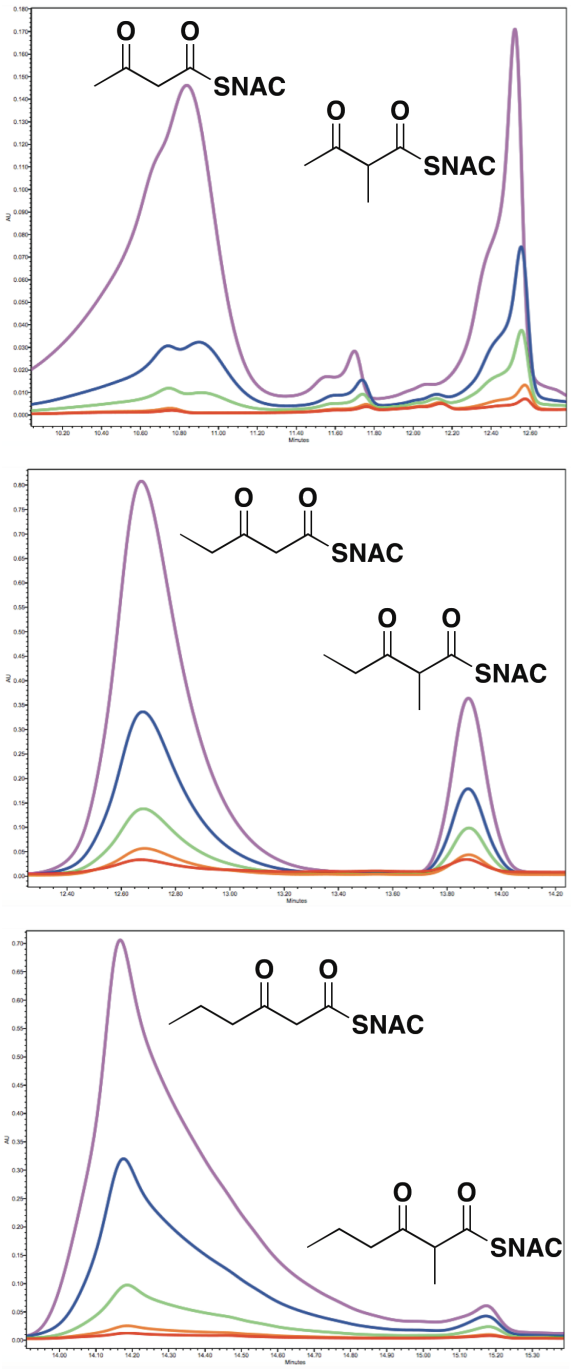


Figure 3-5: MupMT1 methylation activity of 3-oxobutanoyl-*S*-NAC.

Figure 3-5 continued;

HPLC traces monitoring 235 nm for, Top set: MupMT1 catalyzed methylation of 3-oxobutanoyl-*S*-NAC (**1)** to afford 2-methyl-3-oxobutanoyl-*S*-NAC **(4)** at 500 uM (red) 1 mM (orange), 2.5 mM (green), 5 mM (blue), 10 mM (violet) at 1h; Middle set: MupMT1 catalyzed methylation of 3-oxopentanoyl-*S*-NAC (**2)** to afford 2-methyl-3-oxopentanoyl-*S*-NAC **(5)** at 500 uM (red)1 mM (orange), 2.5 mM (green), 5 mM (blue), 10 mM (violet) at 2.5h and Bottom set: MupMT1 catalyzed methylation of 3-oxohexanoyl-*S*-NAC (**3)** to afford 2-methyl-3-oxohexanoyl-*S*-NAC (**6)** at 500 uM (red) 1mM(orange), 2.5 mM (green), 5 mM (blue), 10 mM (violet) at 3h.

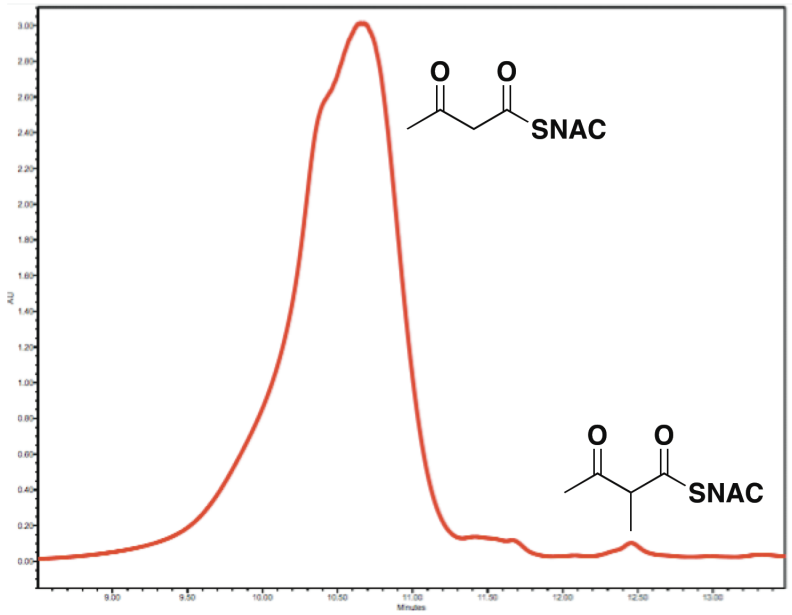


Figure 3-6: HPLC of DifMT1 conversion of 3-oxobutanoyl-*S*-NAC.

HPLC trace monitoring 235 nm of DifMT1 conversion of 3-oxobutanoyl-*S*-NAC (**1**) to 2-methyl-3-oxobutanoyl-*S*-NAC (**4**).

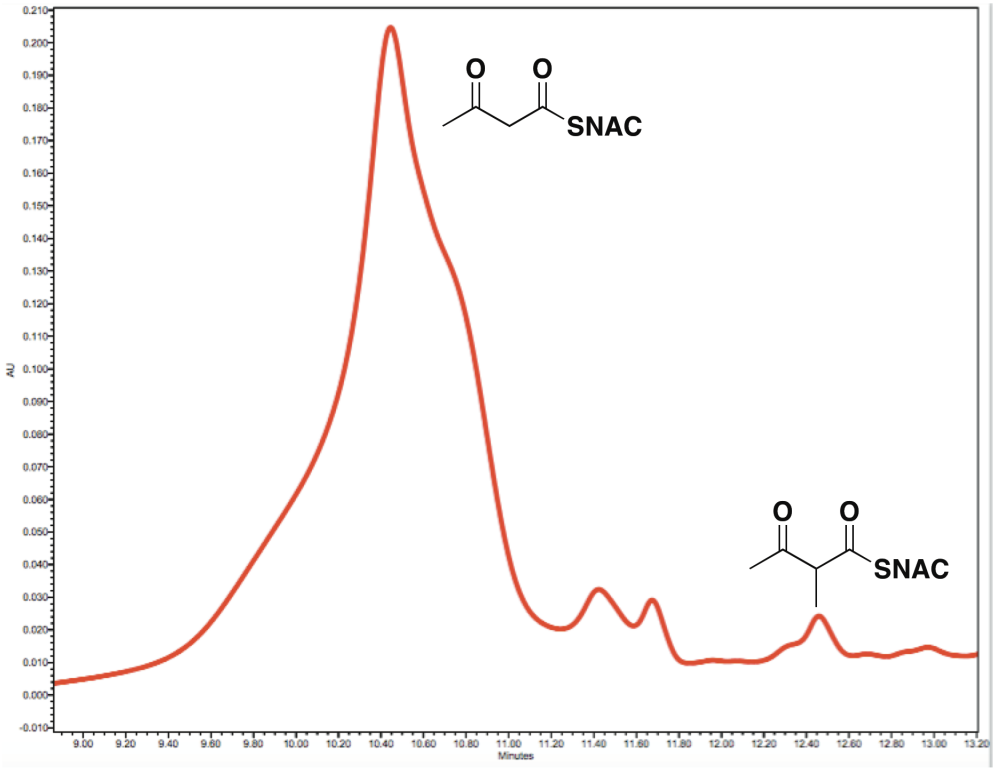


Figure 3-7: HPLC of BaeMT8 conversion of 3-oxobutanoyl-S-NAC.

HPLC trace monitoring 235 nm of BaeMT8 conversion of 3-oxobutanoyl-*S*-NAC (**1**) to 2-methyl-3-oxobutanoyl-*S*-NAC (**4**).

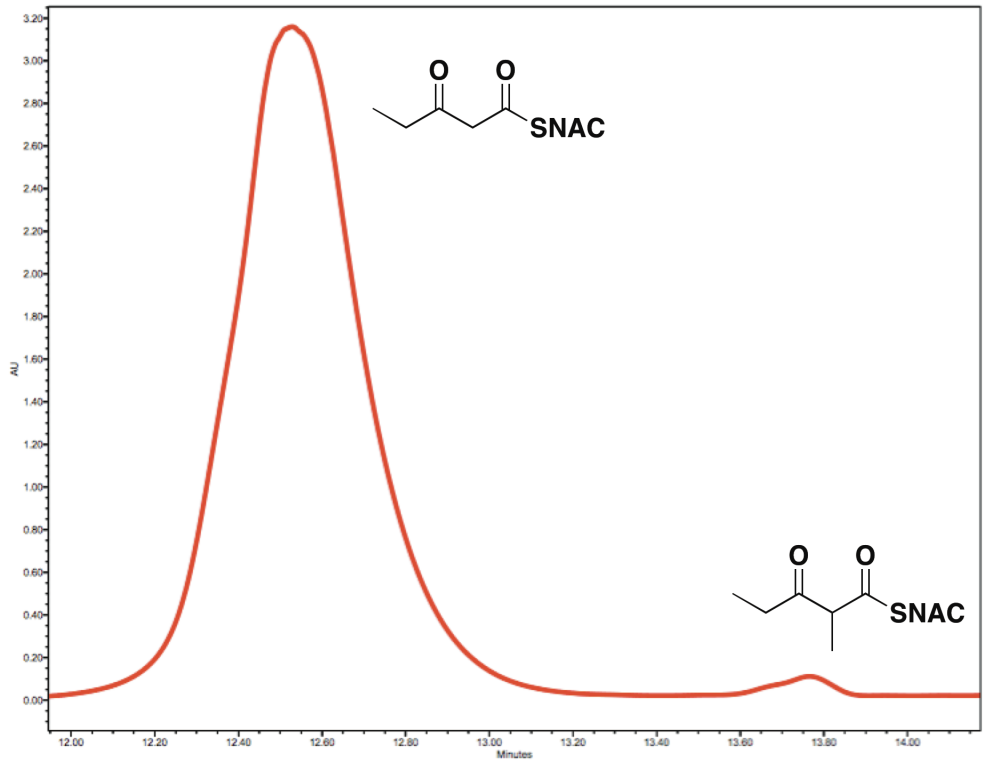
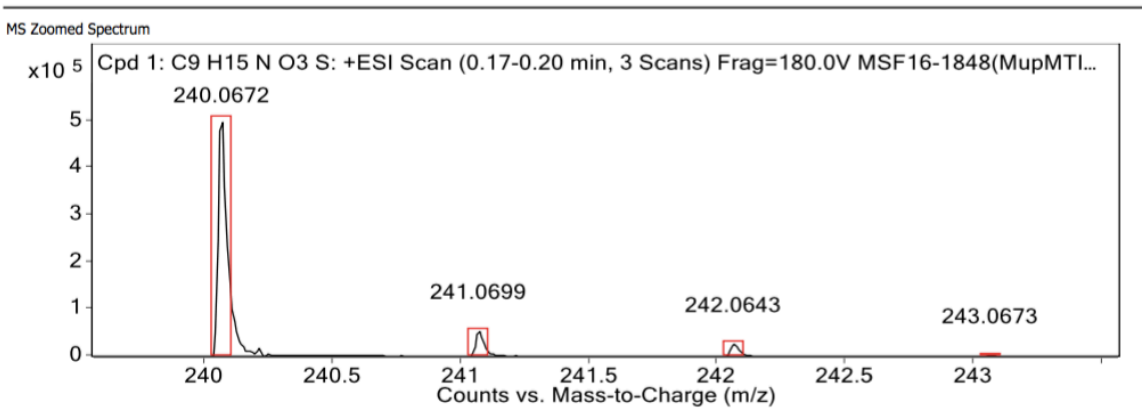


Figure 3-8: HPLC of BaeMT8 conversion of 3-oxopentanoyl-S-NAC.

HPLC trace monitoring 235 nm of BaeMT8 conversion of 3-oxopentanoyl-*S*-NAC (**2**) to 2-methyl-3-oxopentanoyl-*S*-NAC (**4**).



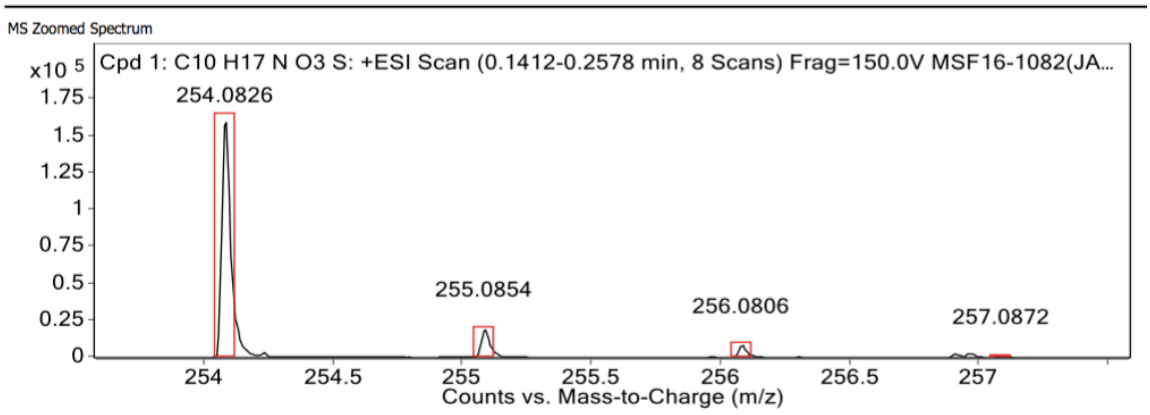
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
240.06720	240.06650	1	507251.1	C9H15NO3S	(M+Na)+	-2.87
241.06990	241.06940	1	54305.59	C9H15NO3S	(M+Na)+	-2.05
242.06430	242.06420	1	26497.82	C9H15NO3S	(M+Na)+	-0.46
243.06730	243.06680	1	2606.54	C9H15NO3S	(M+Na)+	-1.87
244.06660	244.06720	1	3163.47	C9H15NO3S	(M+Na)+	2.45
245.06660	245.06930	1	570.86	C9H15NO3S	(M+Na)+	10.8

--- End Of Report ---

Figure 3-9: HRMS of MupMT1 reaction product 2-methyl-3-oxobutanoyl-*S*-NAC.

HRMS spectrum for MupMT1 produced 2-methyl-3-oxobutanoyl-*S*-NAC (**4**).

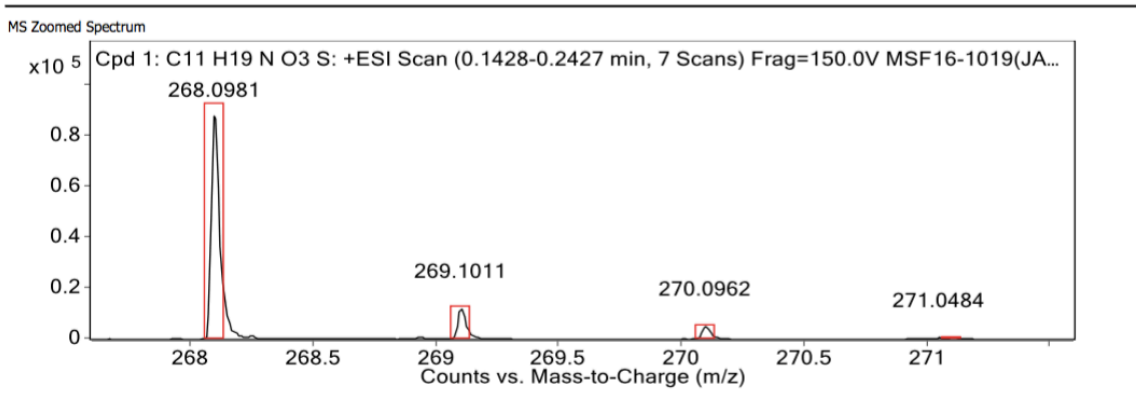


MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
221.00610			303256.5			
254.08260	254.08210	1	164542.14	C10H17NO3S	(M+Na)+	-1.72
255.08540	255.08510	1	19747.87	C10H17NO3S	(M+Na)+	-1.3
256.08060	256.08000	1	8464.31	C10H17NO3S	(M+Na)+	-2.26
257.08720	257.08250	1	1158.46	C10H17NO3S	(M+Na)+	-18.23

Figure 3-10: HRMS of MupMT1 reaction product 2-methyl-3-oxopentanoyl-*S*-NAC.

HRMS spectrum for MupMT1 produced 2-methyl-3-oxopentanoyl-*S*-NAC (**5**).



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
221.00610			285299.58			
268.09810	268.09780	1	92549.85	C11H19NO3S	(M+Na)+	-1.18
269.10110	269.10080	1	12269.99	C11H19NO3S	(M+Na)+	-1.22
270.09620	270.09590	1	5297.56	C11H19NO3S	(M+Na)+	-1.25
271.04840	271.09830	1	747.5	C11H19NO3S	(M+Na)+	184.13

Figure 3-11: HRMS of MupMT1 reaction product 2-methyl-3-oxohexanoyl-*S*-NAC.

HRMS spectrum for MupMT1 produced 2-methyl-3-oxohexanoyl-*S*-NAC (**6**).

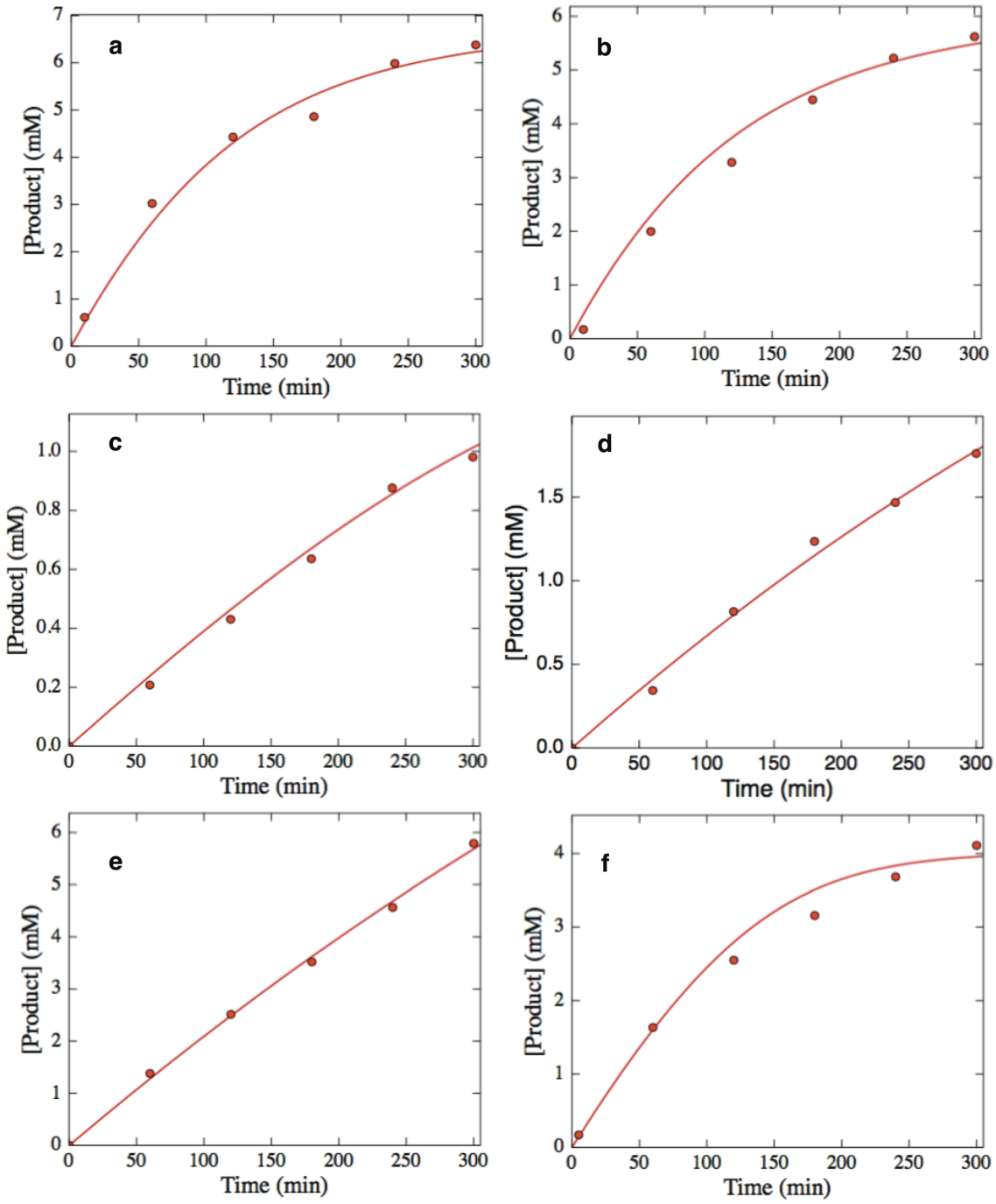


Figure 3-12: Methyltransferase kinetic time course reactions.

Figure 3-12 continued:

Time course data at 10 mM substrate for a) MupMT1 and **1**, b) MupMT1 and **2**, c) MupMT1 and **3**, d) BaeMT8 and **3**, e) DifMT1 and **2**, f) DifMT1 and **3**. Experimental data from concentration time-course was globally fit to a saturation-curve at a fixed time for enzyme-substrate pairs using KinTek Global Kinetic Explorer.

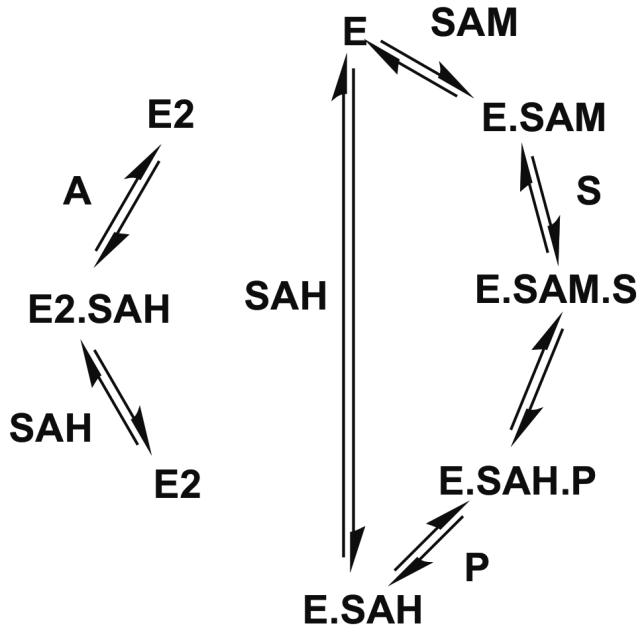
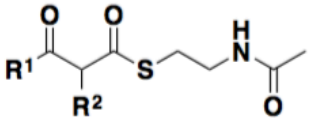


Figure 3-13: Kinetic model for MT activity.

The kinetic model for *in vitro* methyltransferase reactions used to globally fit the data for all experiments using KinTek Global Kinetic Explorer. E=MT; E2=Pfs; SAM=*S*-adenosylmethionine; SAH=*S*-adenosylhomocysteine; S=substrate; P=product; A=adenine.



1, R^1=Me; R^2=H
2, R^1=Et; R^2=H
3, R^1=Pr; R^2=H
4, R^1=Me; R^2=Me
5, R^1=Et; R^2=Me
6, R^1=Pr; R^2=Me

MT domain	substrate	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} mM^{-1})
MupMT1	1	25.8±3.59	1.18±0.326	21.9±6.77
MupMT1	2	46.9±13.5	0.988±0.403	47.5±23.7
MupMT1	3	3.70±1.30	1.33±0.950	2.78±0.796
DifMT1	2	5.27±1.58	0.049±0.017	107.6±49.3
DifMT1	3	4.59±0.76	0.863±0.44	5.32±2.85
BaeMT8	3	2.15±0.965	0.096±0.066	1.22±1.0

Table 3-1: Kinetic analysis of DifMT1, MupMT1, and BaeMT8.

Amplicon	Forward Primer (5' -> 3')	Reverse Primer (5' → 3')
BaeMT8	GCGGCCTGGTGCCGCGCGGCTCTAGCTCACAACAGCCGAAGGTC	GTGGTGGTGGTGGTGGTGATGTTACGCATGTGCTGCTTTTGG
BaeMT12	GCGGCCTGGTGCCGCGCGGCTCTAGCTTGCAAAACACTGAAGATCTTCTTTACAAAC	GTGGTGGTGGTGGTGGTGATGTTAGGAAGCCGCAGTTTCTTCAAC
DifMT1	GCGGCCTGGTGCCGCGCGGCTCTAGCTCCGCACTCGGACAATTAACG	GTGGTGGTGGTGGTGGTGATGTTACTCCGGCTTCTCTTCTGACGC
DifMT6	GCGGCCTGGTGCCGCGCGGCTCTAGCATGAAAGAATTGAATCGGCTTTTGC	GTGGTGGTGGTGGTGGTGATGTTATGTTTGCGGTACGTTTGATTGG
DifMT13	GCGGCCTGGTGCCGCGCGGCTCTAGCATGAACAGACTTCTTTGCAAAC	GTGGTGGTGGTGGTGGTGATGTTACTTTCGGAAAACTTCGGC
MupMT1	GCGGCCTGGTGCCGCGCGGCTCTAGCATGGCAAACATGCTCGCG	GTGGTGGTGGTGGTGGTGATGTTAACTCTCGGCCGAAGGTGC
MupMT3	GCGGCCTGGTGCCGCGCGGCTCTAGCGCTGCGGCCTTGCAAG	GTGGTGGTGGTGGTGGTGATGTTAGGCCAAAGCCACCTTCG
PFS	GCGGCCTGGTGCCGCGCGGCTCTAGCATGAAAATCGGCATCATTGGTG	GTGGTGGTGGTGGTGGTGATGTTAGCCATGTGCAAGTTTCTGCAC

Table 3-2: Primers used for MT plasmid construction.

Primers used for PCR amplification of all methyltransferases assayed as well as the SAH nucleosidase PFS. Ligation independent cloning regions highlighted in red.

Chapter 4: Structural and Functional Trends in Dehydrating Bimodules from *trans*-Acyltransferase Polyketide Synthases

Adapted from Wagner, et al. (2017) *Structure*. **25**, 1045-1055.

ABSTRACT

In an effort to uncover the structural motifs and biosynthetic logic of the relatively uncharacterized *trans*-acyltransferase modular polyketide synthases (*trans*-AT PKSs), we have begun the dissection of the enigmatic dehydrating bimodules common in these enzymatic assembly lines. Here, we report the 1.98 Å resolution structure of a ketoreductase (KR) from the first half of a type B dehydrating bimodule and the 2.22 Å resolution structure of a dehydratase (DH) from the second half of a type A dehydrating bimodule. The KR, from the third module of the bacillaene synthase, possesses features not observed in other structurally-characterized KRs, including a longer active site loop preceding the catalytic tyrosine, differences in the NADPH-binding motif, and a helix that protrudes from the catalytic subdomain. The DH, from the tenth module of the difficidin synthase and the first DH to be structurally elucidated from a *trans*-AT assembly line, possesses residues that may enable α/β-*cis*, γ/δ-*trans* double dehydration as well as a large hydrophobic pocket on its surface that may mediate polypeptide docking for dehydrating bimodules that are split between two polypeptides. Correlations between the chemistries proposed for dehydrating bimodules and bioinformatic analysis indicate that type A dehydrating bimodules generally produce an α/ β -*cis* alkene moiety while type B dehydrating bimodules generally produce a α/β-*trans*, γ/δ-*cis* diene moiety;

however, completely accurate prediction of the intermediates generated by these dehydrating bimodules is not yet possible due to a mysterious isomerase activity.

INTRODUCTION

Modular polyketide synthases (PKSs) are enzymatic assembly lines present in diverse bacterial species that produce bioactive molecules such as the antibiotics erythromycin and mupirocin (Hertweck, 2010). These PKSs have been distinguished as "*cis*-acyltransferase" and "*trans*-acyltransferase" (*cis*-AT and *trans*-AT) PKSs based on the reliance of the *trans*-AT PKSs upon separately-encoded ATs for selecting α-carboxyacyl-CoA polyketide building blocks. While *cis*-AT PKSs, such as the erythromycin PKS, have been fairly well characterized both structurally and functionally, the *trans*-AT PKSs are largely uncharacterized owing to their more recent discovery (Keatinge-Clay, 2012; Piel, 2010; Helfrich and Piel 2016). Beyond the eponymous distinction between these synthase types, *trans*-AT PKSs deviate from the colinearity rules much more often than *cis*-AT PKSs, such that the polyketides produced by *trans*-AT PKSs cannot be confidently predicted by inspecting which enzymes are present within its modules [e.g., ketosynthase (KS), methyltransferase (MT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER)] (Nguyen et al., 2008).

Prominent among the colinearity-breaking domain organizations of *trans*-AT PKSs are the dehydrating bimodules (Figure 4-1). Frequent within *trans*-AT systems, they generally conform to one of two varieties, termed type A and type B (Moldenhauer et al., 2007; Piel, 2010; Helfrich and Piel, 2016). Each dehydrating bimodule type is usually split such that KS and DH of the second module reside in different polypeptides.

Type A dehydrating bimodules are usually correlated with the addition of two carbons in the form of an β-*cis* olefin and are comprised of KS, KR, acyl carrier protein (ACP), and a condensation-incompetent KS ($KS+KR_A+ACP+KS^0$) at the C-terminal end of the upstream polypeptide as well as DH and ACP (DH+ACP) at the N-terminal end of the downstream polypeptide (the KR subscripts indicate an A- or B-type KR, producing an L- or D-β-hydroxyl group, respectively) (Reid et al., 2003; Caffrey, 2003). Type B dehydrating bimodules are usually correlated with the addition of four carbons in the form of an α/β-*trans*, γ/δ-*cis* diene and are comprised of KS, KR, ACP, and KS ($KS+KR_A+ACP+KS$) at the C-terminal end of one polypeptide and DH, ACP, and KR ($DH+ACP+KR_B$) at the N-terminal end of the subsequent polypeptide. If the DH of type B dehydrating bimodules behaves as other characterized DHs, it would be expected to only dehydrate β-hydroxyacyl intermediates and thus operate first on the KR_A-generated L-β-hydroxyacyl intermediate bound to ACP of the first module and then on the KR_B-generated D-β-hydroxyacyl intermediate bound to ACP of the second module. Fascinatingly, these DHs have recently been demonstrated to be capable of catalyzing the double dehydration of D-β-L-δ-dihydroxyacyl-ACP substrates (Challis, G., personal communication).

trans-AT PKSs lack the canonical N- and C-terminal docking domains (NDD and CDD regions) that noncovalently join adjacent PKS polypeptides in *cis*-AT systems (Weissman 2006; Yan, 2009; Whicher, 2013), and the split nature of dehydrating bimodules is representative of the many uncharacterized disconnections between *trans*-AT polypeptides. Recently, the most frequent docking mode for *trans*-AT polypeptides

was elucidated - docking motifs at the C- and N-termini that mediate polypeptide association through the formation of four-helix bundles (Dorival et al., 2016; Zeng et al., 2016). However, the ordered assembly of *trans*-AT polypeptides occurs through other types of docking interactions, such as the noncovalent assembly of split enzymatic domains. The sequences at the polypeptide disconnections of split dehydrating bimodules indicate such a docking mode (Zeng et al., 2016).

Here, crystal structures of enzymatic domains representative of both type A and type B bimodules are presented (Figure 4-1). The 1.98 Å-resolution structure of a KR_A from a type A dehydrating bimodule of the bacillaene PKS (PksKR3) reveals features not observed in other structurally-characterized KRs, including a longer loop preceding the catalytic tyrosine, a shorter lid helix, a different NADPH-binding motif, and a helix that protrudes from the catalytic subdomain (Butcher et al., 2007; Straight et al., 2007). The 2.22 Å-resolution structure of a DH from a type B dehydrating bimodule of the difficidin PKS (DifDH10), the first reported of a *trans*-AT PKS DH, reveals features, such as a histidine that may help catalyze an α/β, γδ-double dehydration as well as a hydrophobic cleft that may play a role in polypeptide docking (Chen et al., 2006). Correlations between the chemistries proposed for dehydrating bimodules and bioinformatic analysis indicate that type A dehydrating bimodules generally produce an α/β-*cis* olefin while type B bimodules generally produce a α/β-*trans*, γ/δ-*cis* diene; however, an uncharacterized isomerase activity still precludes accurate prediction of the intermediates generated by dehydrating bimodules.

RESULTS AND DISCUSSION

Characteristics of dehydrating bimodules

To examine the composition of type A and type B dehydrating bimodules, representatives were aligned from the polypeptides of the bacillaene (abbreviations Bae and Pks represent the assembly lines from the *Bacillus amyloliquefaciens FZB42* and *Bacillus subtilis str*. 168, respectively), batumin, chivosazol, difficidin, elansolid, kirromycin, leinamycin, macrolactin, sorangicin, and thailandamide assembly lines (type A: BaeK/BaeL, BaeL/BaeM, Bat2/Bat3, DifG/DifH, PksK/PksL, PksL/PksM, SorH/SorI, TaiK/TaiL; type B: ChiC, ChiD/ChiE, ChiF, DifI/DifJ, ElaO/ElaP, KirAI/KirAII, KirAIV/KirAV, LnmI/LnmJ, MlnE/MlnF, of these the only dehydrating bimodules not split between polypeptides reside in ChiC and ChiF)(Piel, 2010; Helfrich & Piel, 2016; Butcher et al., 2007; Chen et al., 2006; Mattheus et al., 2010; Perlova et al., 2006; Teta et al., 2010; Schneider et al., 2007; Weber et al., 2008; Nguyen et al., 2008) (Figure 4-2). A third type of dehydrating bimodule (termed "type C" here) was also observed, equivalent to a type A dehydrating bimodule but followed by a DH-like domain that apparently functions as an isomerase interconverting α/β-*cis*, δ/γ-*trans* dienes with α/β-*trans*, δ/γ-*cis* dienes (e.g., DifJ/DifK, BonA/BonB, SorA/SorB, see discussion in *Predicting the Polyketide Intermediates Generated by Dehydrating Bimodules*).

Aside from the presence of KR_Bs at the C-terminal end of type B dehydrating bimodules, type A and type B dehydrating bimodules are structurally quite similar (Figure 4-1). The first KS and the KS usually following the bimodule contain flanking subdomains, while the second KS does not (Gay et al., 2014a). The KR_A present in both

bimodule types belongs to an unusual clade of KRs rarely observed in *trans*-AT PKSs outside of dehydrating bimodules. Both type A and type B dehydrating bimodule DHs lack ~20 residues observed at the N-terminal end of canonical DHs from *trans*-AT and *cis*-AT PKSs.

Several of the domains within type A and type B dehydrating bimodules belong to distinct clades (Figure 4-2). This is not surprising for KSs, since KS clades are strongly correlated with the chemistry performed by the processing enzymes of the previous module (Nguyen et al., 2008). Thus, while the first KS of dehydrating bimodules can belong to several clades, the second KS is from clade XIV or XV (type A or B, respectively) and if a KS follows a dehydrating bimodule it belongs to clade XI or IX (type A or B, respectively). While the KR_As of type A and type B dehydrating bimodules clade together, the DHs belong to slightly different clades, possibly due to differences in their activities and docking interactions. While equivalent ACPs of type A and type B dehydrating bimodules seem to belong to the same clade, the ACPs from the first module of type A and type B dehydrating bimodules belong to a different clade than those from the second module. When tandem ACPs are present, as in ChiC, MlnG/MlnH, and MlnE/MlnF, they clade together. The ACP of the second module of a type B split bimodule precede the KR domain, as does the ACP preceding the DH-like isomerase in type C dehydrating bimodules; this unusual ordering is also observed in *trans*-AT modules containing an embedded ER.

KR_A Structure & Activity

The structure of PksKR3 was solved to 1.98-Å resolution by molecular replacement using PksKR2 as a search model (PDB: 4J1Q)(Piasecki et al., 2014)(Figure 4-2 and Table 4-1). Like previously structurally-characterized KRs from *cis*-AT PKSs (Keatinge-Clay & Stroud, 2006; Keatinge-Clay, 2007; Zheng et al., 2010; Zheng et al., 2013a, Zheng et al., 2013b, Bonnett et al., 2013) and *trans*-AT PKSs (Piasecki et al., 2014; Zeng et al., 2016), PksKR3 is comprised of an N-terminal structural subdomain (KR_s) and a C-terminal catalytic subdomain (KR_c), each possessing a Rossmann fold (Figure 4-3). The KR_c includes a catalytic tyrosine, Y386, as well as the conserved active site lysine and serine residues, K371 and S347 (numbering based on first observable residue). Complete density was observed for co-crystallized $NADP^+$.

Several features in PksKR3 are distinct from other solved *cis*- and *trans*-AT PKS KRs (Figures 4-3a-c, and 4-2). The NADPH binding motif is GGTRG(I/L)G rather than the typical motif GGxGG[I,L]G (Simunovic et al, 2006). The arginine of this motif, R228, makes ionic interactions with both the 2'- and 5'-phosphates of the nicotinamide cofactor, in contrast to an arginine in other solved KRs, located between βB and αC, that only contacts the 2'-phosphate (Figure 4-3c) (Keatinge-Clay, 2006). The glutamine three residues before the catalytic tyrosine (Q383), which in other KRs helps orient the polyketide substrate for stereoselective reduction through hydrogen bonding interactions with its thioester carbonyl, is itself oriented through a hydrogen bond with a carbonyl in a helical turn not present in other solved KRs (Mugnai et al., 2015; Keatinge-Clay, 2016). PksKR3 contains an extra β-strand in KR_s compared to other structurally-determined

KRs as well as a shorter lid helix. Loop DE, which in B-type KRs contains the signature aspartate of B-type KRs, possesses the consensus sequence HCAGxxxxxx[P,L]AF[I,L]xK (residues 320-335) (Keatinge-Clay & Stroud, 2006).

The most striking difference between PksKR3 and other structurally-characterized KRs is an ~15-residue motif that contains the consensus sequence LPPxxxW (residues 255-271, ~25 residues C-terminal to the NADPH binding motif) and forms αBC, which makes very little contact with the rest of PksKR3 (Figures 4-3b and 4-3c). The LPPxxxW motif, conserved in KRs from dehydrating bimodules, is also present in many KRs following the adenylation (A) domains of non-ribosomal peptide synthetase (NRPS) modules and in the KRs from tubulysin modules 4 and 6 (Figure 4-4) (Hur et al., 2012; Sandmann et al., 2004). Interestingly, a C-methyltransferase (CMT) is present immediately C-terminal to this motif in tubulysin module 6. The structure of the CMT embedded in curacin module 7 (PDB: 5THZ) reveals how a partially conserved LPPDFLL motif at its N-terminal end fills in a hydrophobic groove (Skiba et al., 2016). Thus, the LPPxxxW motif could be the vestiges of a CMT that had inserted into a KR domain.

The *cis*-double bond generated by type A dehydrating bimodules indicates that the first KR of type A and type B dehydrating bimodules is an A-type KR; also suggestive is that these KRs lack the signature aspartate of B-type KRs (Moldenhauer et al., 2010; Keatinge-Clay, 2016; Piasecki et al., 2013). To test this, PksKR3 was incubated with --ketopentanoyl-*S*-*N*-acetylcysteamine (NAC) in the presence of an NADPH regeneration system (*Bacillus subtilis* glucose dehydrogenase, D-glucose, and $NADP^+$),

and the products were analyzed by chiral HPLC (Piasecki et al. 2011) (Figure 4-3d). Although NAC-bound substrates are shorter than natural ACP-bound substrates, they are often reduced equivalently to the natural substrate (Piasecki et al., 2011; Bailey et al., 2016). For comparison, identical ketoreduction assays were performed with two well-characterized KRs [the KR from the first module of the tylosin synthase (TylKR1) that generates (*3R*)-hydroxypentanoyl-*S*-NAC, and the KR from the fifth module of the mycolactone synthase (MycKR5) that generates the (*3S*)-hydroxypentanoyl-*S*-NAC in 93% enantiomeric excess (*ee*)] (Piasecki et al., 2011). In agreement with the prediction of PksKR3 as an A-type KR, the major product was (*3S*)-hydroxypentanoyl-*S*-NAC (64% *ee*) (Figure 4-3). The assignment of PksKR3 as an A-type KR renders it the first A-type KR from a *trans*-AT PKS to be structurally characterized. All of the type A, B, and C dehydrating bimodules analyzed here possess this type of A-type KR within their first module.

DH Structure & Activity

The DH from the type B dehydrating bimodule of the difficidin synthase (DifDH10), was expressed, purified, and crystallized (Figures 4-5 and Table 4-1). Its structure was solved to 2.22-Å resolution by molecular replacement using a polyalanine model of the DH from the tenth module of the rifamycin synthase as a search model (PDB: 4LN9) (Gay et al., 2013). DifDH10 possesses a double hotdog fold similar to a recently-reported DH from a type A dehydrating bimodule (PksDH4, PDB: 5E1V), DHs from *cis*-AT PKSs (PDB: 3EL6, 3KG6, 3KG7, 3KG8, 3KG9), as well as the closely-related branching (B) domain and enoyl-isomerase (PksEI16) (PDB: 4KC5,

4U3V)(Keatinge-Clay, 2008; Akey et al., 2010; Bretschneider et al., 2013; Gay et al., 2014). Unlike the dimeric DHs of *cis*-AT PKSs and the B domain, but like PksDH4 and PksEI16, DifDH10 crystallized as a monomer.

When the DHs from split type A and type B dehydrating bimodules (PksDH4 and DifDH10) are compared with related domains, several differences are evident (Figures 3a-e). The most striking deviation is that they possess ~20 fewer N-terminal residues. The HPLL motif present in DHs from both *cis*- and *trans*-AT PKSs occupies a hydrophobic pocket adjacent to α1 (Keatinge-Clay, 2008); with these residues absent, the hydrophobic pocket is exposed (Figures 4-5a and 4-5b). If the N-terminal DHs of split dehydrating bimodules mediate the association with the upstream polypeptide, then the residues at the C-termini of these polypeptides could complement the hydrophobic pocket. Indeed, semi-conserved residues are present following the C-terminal KSs of the first polypeptide of split type A and type B dehydrating bimodules and could enable these KSs to associate with N-terminal DHs to form a structure similar to the KS+B homodimer that was structurally elucidated from the rhizoxin assembly line (PDB: 4KC5) (Figures 4-5a, 4-5b, and 4-7)(Bretschneider et al., 2013).

The ~20-residue linker connecting the two halves of the DH double hotdog differs between the DHs of *cis*-AT and *trans*-AT assembly lines. The last portion of this sequence in the DHs of *cis*-AT assembly lines contains a conserved WPP motif that is replaced by a helical motif in the DHs of *trans*-AT PKSs (residues 126-134 in DifDH10). The first, third, and sixth residues of this motif are large hydrophobes while the second residue is most often an aspartate or asparagine that caps the helix (LDLSQIKAS in

PksDH4 and LNIEAFMKN in DifDH10). As with the WPP motif, these hydrophobic residues contact a hydrophobic patch on the DH β-sheet.

The active sites of PksDH4 and DifDH10 resemble all other structurally-determined DHs with their catalytic histidine and aspartic acid equivalently oriented (H25 and D198 in DifDH10, numbering based on first observable residue) (Figure 4-5). However, the tyrosine of *cis*-AT DHs hypothesized to form a hydrogen bond with the β-hydroxyl group of a substrate prior to dehydration (in a GYx**Y**GPxF motif) is replaced by a histidine (termed H' here, H155 is 42 residues before the catalytic aspartic acid in DifDH10) and serves as a signature of DHs from dehydrating bimodules (Figure 3d) (Keatinge-Clay, 2008; Keatinge-Clay, 2016). H' seems to assume the role of the histidine or glutamine four residues after the catalytic aspartic acid in DHs from *cis*-AT assembly lines, forming a hydrogen bond with the aspartic acid (Figure 4-5c). Dehydration mediated by the DHs of type A dehydrating bimodules may proceed how *cis*-AT DHs are hypothesized to convert an L-β-hydroxyacyl substrate to a *cis*-α/β double bond-containing product, with the aspartic acid donating a proton to the β-hydroxy leaving group and the histidine abstracting a proton from the L-α-hydrogen (Figure 3a) (Valenzano et al., 2010; Keatinge-Clay, 2016). The double dehydration mediated by the DHs of type B dehydrating bimodules on D-β-L-δ-dihydroxyacyl substrates is likely to first proceed how *cis*-AT DHs are hypothesized to convert a D-β-hydroxyacyl substrate to a *trans*-α/β double bond-containing product (Figure 4-5b) (Keatinge-Clay, 2016). How the second dehydration occurs is less certain. We propose that the thioester shifts towards the surface of the protein to form a hydrogen bond with a histidine conserved in DHs

from type B dehydrating bimodules (termed H" here, in the Hxx**H**GxxxxP motif); this would facilitate the abstraction of a proton from the L-γ-hydrogen by the same histidine that abstracted the proton from the L-α-hydrogen (**H**xxHGxxxxP) as the aspartic acid donates a proton to the L-δ-hydroxy leaving group to enable the second dehydration. To test the dehydrating activity of DifDH10, the substrate mimics crotonyl-*S*-NAC and (*E*)-hex-2-enyl-*S*-pantetheine were assayed. Hydration of both compounds was observed (Figure 4-6).

Predicting Dehydrating Bimodule Polyketide Intermediates

Type A split bimodules usually produce α/β-*cis* olefins, and type B split bimodules usually produce α/β-*trans*, γδ-*cis* dienes. Some biosynthetic models for assembly lines containing split bimodules should be updated to reflect this. The type B dehydrating bimodules formed by DifI/DifJ and MlnE/MlnF from the difficidin and macrolactin assembly lines should each be shown producing an α/β-*trans*, γ/δ-*cis* diene. The biosynthetic model for the kirromycin assembly line is more challenging, since the α/β-*trans*, γ/δ-*cis* diene likely generated by the dehydrating bimodule formed by KirAIV/KirAV cannot be corroborated by the structure of kirromycin due to a downstream oxidation of the γ/δ-olefin (Weber et al, 2008). What most thwarts the accurate predifction of the polyketide intermediates, however, are mysterious isomerizations (Figure 4-7)

The biosynthetic models for the kalimanticin (equivalent to batumin) and thailandamide assembly lines should seemingly be updated such that the type A dehydrating bimodule in each of them is shown generating a *cis*- rather than a *trans*-

double bond as kalimanticin A and thailandamide B each contain a *cis*-double bond in the anticipated positions (Mattheus et al., 2010; Nguyen et al., 2008). However, kalimanticin B and thailandamide A are generated by the same assembly lines and each contain a *trans*-double bond in these positions. The *trans*-double bonds of kalimanticin B and thailandamide A may be generated through uncatalyzed *cis-trans* isomerization of kalimanticin A and thailandamide B; however, as discussed below, similar isomerizations from type B dehydrating bimodules actually appear to be catalyzed during polyketide synthesis.

The biosynthetic models for the macrolactin and chivosazol assembly lines seemingly need to be updated such that the type B dehydrating bimodules formed by MlnG/MlnH and ChiD/ChiE are shown generating α/β-*trans*, γδ-*cis* dienes rather than α/β-*cis*, γδ-*trans* dienes (Schneider et al., 2007; Perlova et al., 2006). However, this would be inconsistent with the double bond geometries present within macrolactin and chivosazol. The type B dehydrating bimodule in the macrolactin assembly line is immediately followed by the macrocyclizing thioesterase (TE) that offloads the fully constructed polyketide, thus the type B dehydrating bimodule is apparently capable of generating an α/β-*cis*, γ/δ-*trans* diene. Perhaps the DH in this type B dehydrating bimodule is capable of interconverting an α/β-*trans*, γ/δ-*cis* diene with an α/β-*cis*, γ/δ-*trans* diene.

Consistent with this, DH-like isomerases seem to interconvert α/β-*trans*, γ/δ-*cis* dienes with α/β-*cis*, γ/δ-*trans* dienes in type C dehydrating bimodules. Examples appear in the assembly lines that produce bongkrekic acid, difficidin, and sorangicin (bimodules

formed by BonA/BonB, DifJ/DifK, and SorA/SorB) (Moebius et al., 2012; Chen et al., 2006; Irschik et al., 2010). In each case an α/β-*trans* double-bonded intermediate (from *trans*-double bond-generating modules in the bongkrekic acid and sorangicin assembly lines and from a type B dehydrating bimodule in the difficidin assembly line) is passed to a type C dehydrating bimodule. After the installation of a *cis*-double bond by the type C dehydrating bimodule, the α/β-*cis*, γ/δ-*trans* diene is apparently interconverted with an α/β-*trans*, γ/δ-*cis* diene by the C-terminal DH-like enzyme. A downstream enzyme such as KS may select the isomerized intermediate.

Insights into Assembly Lines from Dehydrating Bimodules

This study of dehydrating bimodules raises more questions about *trans*-AT assembly lines than it answers. Type A dehydrating bimodules contain two more domains than a functionally equivalent module. Why dehydrating bimodules appear in place of canonical modules is unclear. That the vast majority of dehydrating bimodules are split between polypeptides even though they can exist on one polypeptide (e.g., ChiC, ChiF) indicates that the split confers an advantage, perhaps in terms of folding or evolution. Thus, this split may confer at least split dehydrating bimodules an advantage over canonical modules.

The mode through which the polypeptides associate is informative. The current data support that the C-termini of the upstream polypeptides docks in the hydrophobic DH pocket filled by the HPLL motif in canonical DHs (Challis, personal communication). Split DHs have been observed to mediate polypeptide assembly in *cis*-AT assembly lines, although the split is between the two hotdog subdomains (Amagai et

al., 2013). Essentially unique within *trans*-AT assembly lines to the first module of dehydrating bimodules, KR_A could also play a role in polypeptide docking. If KR_A assumed an orientation as shown in the model (Figure 4-8), the LPPxxxW motif and αBC would be close enough to contact the downstream, N-terminal DH. The study of the several modes through which *trans*-AT polypeptides dock is only in its infancy (Dorival et al., 2016; Zeng et al., 2016).

CONCLUSIONS

This study highlights the versatility of DH-like enzymes. The DHs of type A dehydrating bimodules apparently mediate two dehydrations to generate both *trans*- and *cis*-double bonds. They may also mediate the interconversion of α/β-*trans*, γ/δ-*cis* and α/β-*cis*, γ/δ-*trans* dienes as appears to be the role of the DH-like enzyme at the C-terminal end of type C dehydrating bimodules. Other DH-like enzymes mediate double bond-shifting, either in the same site that catalyzed a dehydration or in an active site devoted to mediating the isomerization, as in the case of enoyl-isomerases (Moldenhauer et al., 2010; Gay et al., 2014). DH-like enzymes can even mediate pyran formation (Pöplau et al, 2013). The diverse chemistries catalyzed by these highly-related enzymes warrant mechanistic study.

The KSs that follow the first and second modules of dehydrating bimodules belong to specific clades implicated in gatekeeping to ensure only specified chemistries are passed down the assembly line (Nguyen et al., 2008). The TE in the macrolactine assembly line may also be performing a gatekeeping function, only performing the cyclization of the α/β-*cis*, γ/δ-*trans* isomer. Studies to understand the mechanisms of

gatekeeping and the range of intermediates from which the specified intermediate is selected will further reveal the functional details of *trans*-AT assembly lines.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification

PksKR3 was amplified from *B. subtilis str. 168* genomic DNA with primers: 5'-*GCGGCCTGGTGCCGCGCGGCTCTAGC*TCAGAAAGAGACAAAAAAGAACTG -3' and 5'- *GTGGTGGTGGTGGTGGTGATGTTA*TCCATCACATTTGGAAACTTC-3' (homology region for ligation independent cloning indicated in italics) and cloned into expression plasmid pGAY28b (Gay et al., 2014c). The construct encoded for residues 3954 – 4459 of the PksK protein, with domain boundaries chosen based on previously solved KR structures. DifDH10 was amplified from *Bacillus amloliquefaciens FZB42* genomic DNA with primers: 5'-AAGGTTCAA*CATATG*GCAGTTGCCACCGATCAAGTCAAAACG -3'
and 5'- TTGAACCTT*CTCGAG*ATGGGTAATCATATTCGCTTGTCTGACCCT - 3' (NdeI and XhoI restriction sites indicated in italics). Domain boundaries were determined via comparing DifDH10 to EryDH4 (Keatinge-Clay, 2008). The amplicon was then digested with NdeI and XhoI and ligated into pET21a. The plasmids were then transformed into E. coli Bl21(DE3).

For protein expression both for PksKR3 and DifDH10, *E. coli* BL21(DE3) cells transformed with the appropriate plasmid were grown to an OD_{600} of 0.5 in Luria broth containing 50 mg/L kanamycin at 37°C. The temperature was then dropped to 15°C prior to inducing protein expression with 0.5 mM IPTG, and grown for an additional 16 h. For

PksKR3, cells were collected via centrifugation at 4,000 x g for 15-30 minutes, resuspended in lysis buffer (400 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol), and lysed by sonication. Cellular debris was removed by centrifugation at 30,000 x g for 30 minutes and the cell lysate was then poured over Ni-NTA resin (Qiagen) pre-equilibrated with lysis buffer. Bound protein was washed in 40 ml of lysis buffer containing 15 mM imidazole and eluted with lysis buffer containing 150 mM imidazole. An equilibrated Superdex 200 gel filtration column was used to buffer exchange the protein solution into 150 mM NaCl and 10 mM HEPES pH 7.5. PksKR3 was concentrated to 11 mg/mL prior to crystallization trials. DifDH10 was purified identically with the exclusion of glycerol from the lysis, elution, and wash buffers. DifDH10. Prior to crystallization trials, DifDH10 was concentrated to 15 mg/mL.

Crystallization and Structure Determination

Crystals of PksKR3 grew in 2 days by sitting drop vapor diffusion at 22 °C. Drops were formed by mixing 2 μL protein solution (9 mg/mL PksKR3, 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 μL crystallization buffer (sodium citrate, 0.1 M HEPES pH 7.0). Crystals were frozen in liquid nitrogen after a 20-min soak in the crystallization buffer modified with 10% (v/v) ethylene glycol. Diffraction data, collected at ALS Beamline 5.0.3, were processed by HKL2000 (Otwinowski & Minor, 1997). The structure was solved to 1.98 Å resolution by molecular replacement with PhaserMR (McCoy et al., 2007) in the CCP4 suite (Winn et al., 2011), using the KR1 monomer from the PksX synthase of *B. subtilis* (PDB code: 4J1Q) as the search model. The model generated from the molecular replacement solution was used to iteratively build into the

remaining electron density map with Coot (Emsley et al., 2010) and was refined with Refmac5 (Murshudov et al., 1997). The only regions that could not be modeled were residues 69-73 and 207-210. The first 14 and final 33 residues encoded in the PksKR3 construct were absent in the x-ray structure.

Crystals of DifDH10 were grown by the sitting vapor diffusion for 2 days at 22 °C, in a equilibration solution of a 2:1 mix of protein solution (~15 mg/mL) and crystalization solution [26% PEG4000, 0.2 M lithium sulfate, 0.1 M Tris-HCl, pH 8.5]. Crystals were briefly soaked in cryoprotectant with 10% (v/v) glycerol added in the crystallization buffer and frozen in liquid nitrogen. The diffraction data, collected at the GM/CA beamline 23-ID-B at the Advanced Photon Source (APS, Argonne National Laboratory), were processed with iMosflm in the CCP4 suite. The structure was determined to 2.22 Å resolution by molecular replacement with PhaserMR using a polyalanine model of RifDH10 as the search model (PDB: 4LN9) (Gay et. al, 2013). The structure was solved in a primitive triclinic system with four chains in the symmetric unit. The model generated from the molecular replacement solution was refined with Coot and Refmac5.

Substrate Synthesis

3-ketopentanoyl-*S*-NAC: This KR substrate was synthesized as previously reported (Piasecki et al., 2011). (E)-hex-2-enyl-*S*-pantetheine: A solution of D-pantetheine (90 mg, 0.35 mmol, 1 eq.) dissolved in 5 mL dichloromethane was cooled to 4° C. To this solution was added (*E*)-hex-2-enoic acid (40 mg, 0.35 mmol, 1 eq.), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (67.1 mg, 0.35 mmol, 1 eq.), and

dimethylaminopyridine (4 mg, 0.035 mmol, 0.1 eq.) and was stirred at 4 °C for 1 h and at 22 °C for another hour. Solvent was removed *in vacuo*, and the remaining material then purified via flash column chromatography (silica, 1:1 EtOAc:acetone) to afford (*E*)-hex-2-enyl-*S*-pantetheine (100 mg, 81%) as a clear oil. The characterization was in agreement with a prior synthesis (Gay et al, 2013).

DifDH10 Hydration Assay

Hydration reactions were performed with 0.5 mM crotonyl-pantetheine, 0.25 M NaCl, and 150 μM DifDH10 in 25 mM HEPES buffer (pH 7.5) at a total volumne of 100 uL at room temperature for 4 h. As a positive control, identical reactions were performed with 150 μM RifDH10 (Gay et al. 2013). Reactions were quenched by adding 50 μL of methanol. The reactions were then centrifuged at 20,000 x g for 8 min to remove precipitated protein. The samples were then analyzed on an Agilent 1200 Series HPLC coupled to an Agilent Technologies 6130 quadrupole mass spectrometer with an electrospray-ionization (ESI) source. A gradient of acetonitrile/water with 0.1% formic acid (5% - 95%) was run on an C_{18} reversed-phase analytical column (5 μm, 2 × 50 mm; Phenomenex) over 12 min at a flow rate of 0.7 mL/min and monitored at 230 nm.

PksKR3 Reduction Assay

Ketoreduction reactions were modified from a method described previously (Piasecki et al. 2011), with 5 mM 3-ketopentanoyl-*S*-NAC, 125 mM HEPES pH 7.5, 250 mM NaCl, 200 mM D-glucose, 10% v/v glycerol, 200 μM $NADP^+$, 1 μM glucose dehydrogenase (from *B. subtilis str*. 168), and 5 μM ketoreductase in a total volume of

500 μL The reactions were incubated overnight at room temperature (25 °C), extracted with ethyl acetate (3 x 500 μL) and dried in a speedvac. All samples were resuspended in ethanol prior to chromatographic analysis. Chiral chromatography was performed with a ChiraCel OC-H column (250 x 4.6 mm) with a Beckman Coulter System Gold 126 pump and a System Gold 166 PDA detector equipped with a 20μL loop. Absorbance was monitored at 235 nm. The solvent system used was 7% ethanol in hexanes (measured using volumetric flasks) at 0.8 mL/min. Enantiomeric excess was determined using peak area integrations.

Accession Numbers

Coordinates for PksKR3 and DifDH10 have been deposited in the PDB under accession codes 5KTK and 5KKU, respectively.

ACKNOWLEDGEMENTS

Instrumentation and technical assistance for crystallographic work were provided by Dr. Art Monzingo and the Macromolecular Crystallography Facility, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology at the University of Texas at Austin. The Advanced Light Source (ALS) is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231. The Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy Office of Science by Argonne National Laboratory, was supported by the US DOE under contract number DE-AC02-

06CH11357. We thank the NIH (GM106112) and the Welch Foundation (F-1712) for supporting this research (A.T.K.)

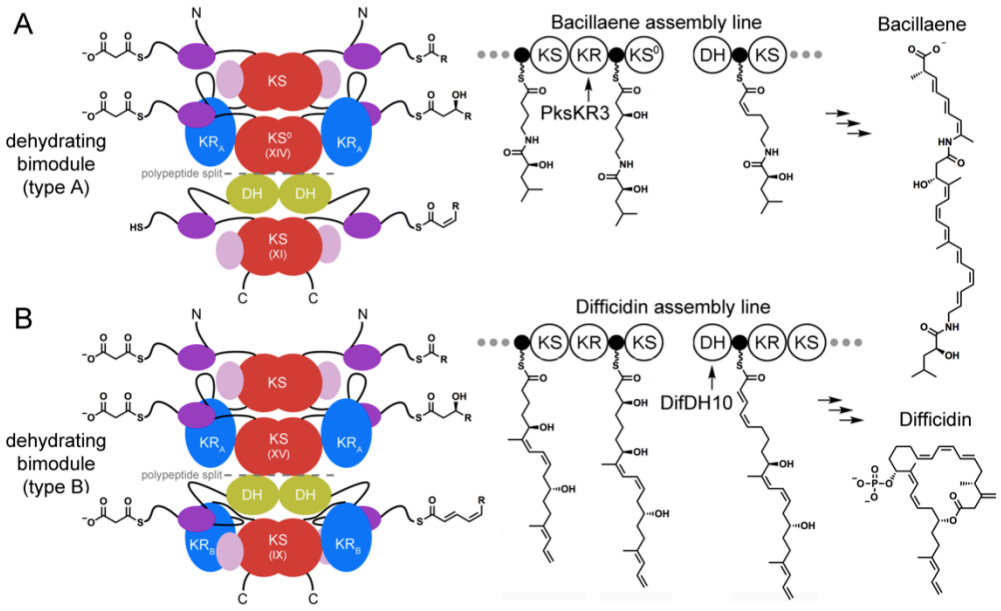


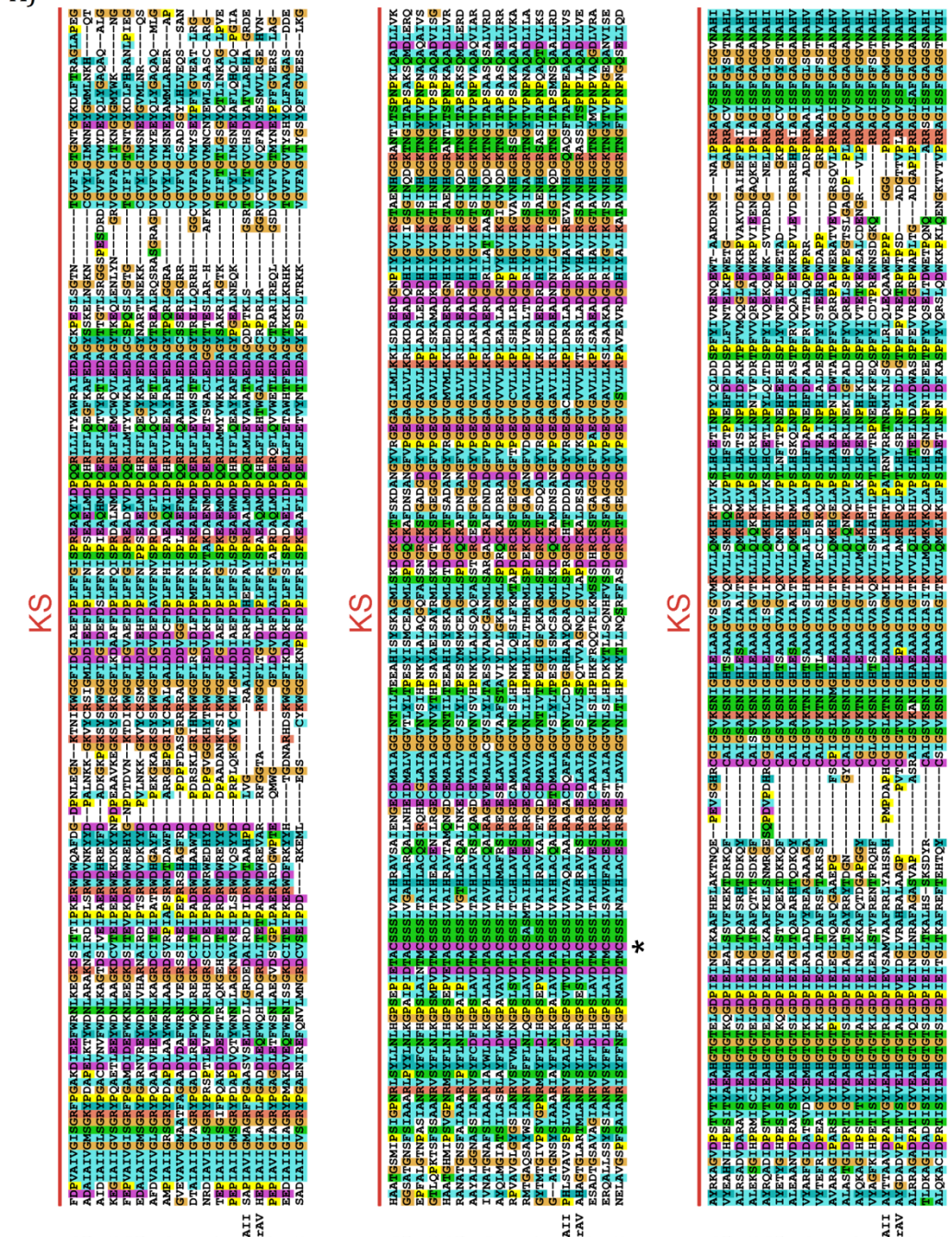
Figure 4-1: Schematics of Type A and B Dehydrating Bimodules.

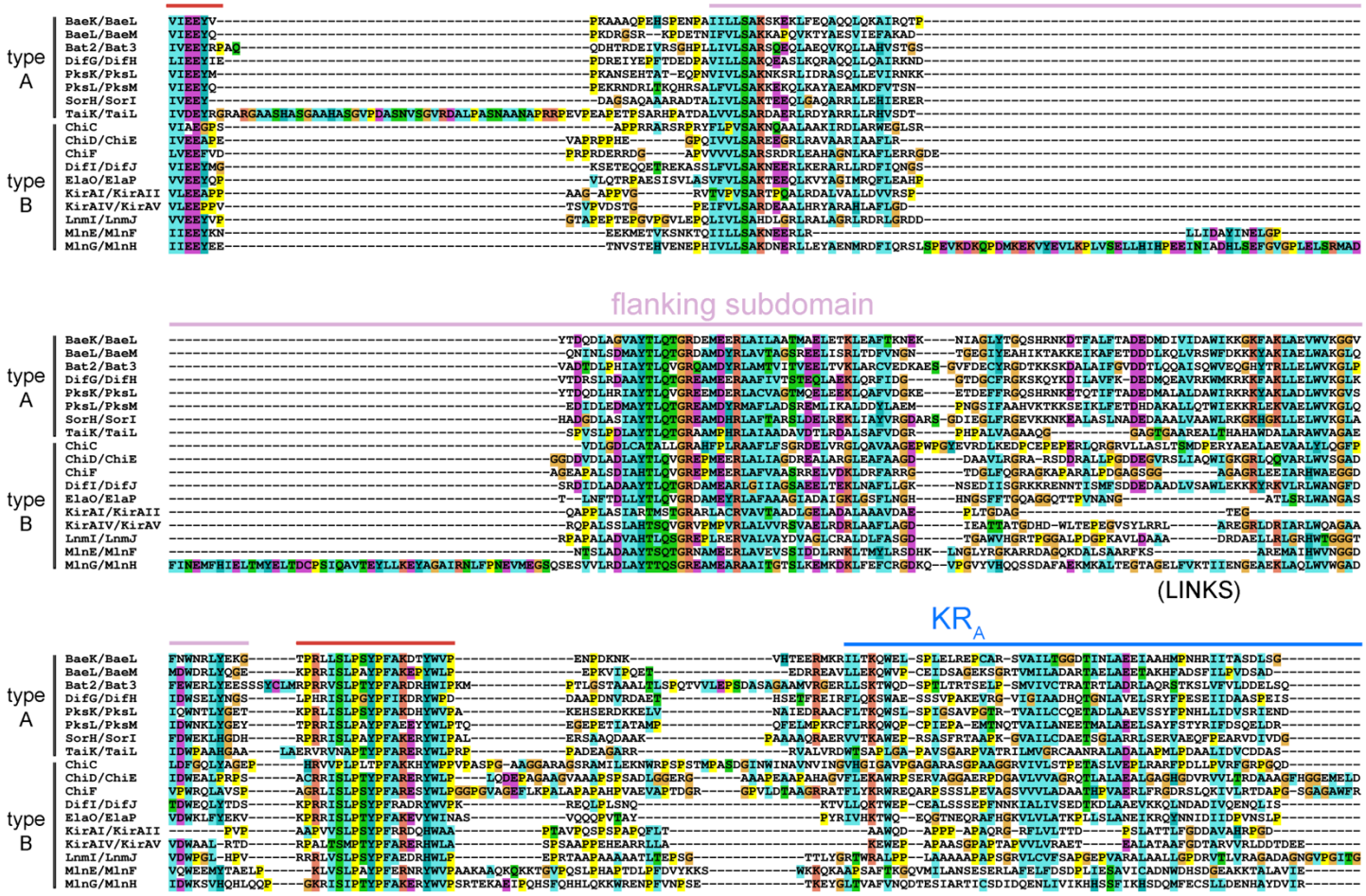
(A) Type A dehydrating bimodules are typically split into polypeptides containing KS + KR_A + ACP + KS_0 and DH + ACP. The KS that follows the first module belongs to clade XIV, and a KS that follows the second module belongs to clade XI (Nguyen et al., 2008). The left side of the homodimeric assembly line shows which ACPs extend the growing polyketide, while the right side shows the hypothesized chemistry of the intermediates as they pass through the bimodule. PksKR3 is the KRA within the first module of the type A dehydrating bimodule formed by the polypeptides PksK/PksL from the bacillaene assembly line of B. subtilis 168.

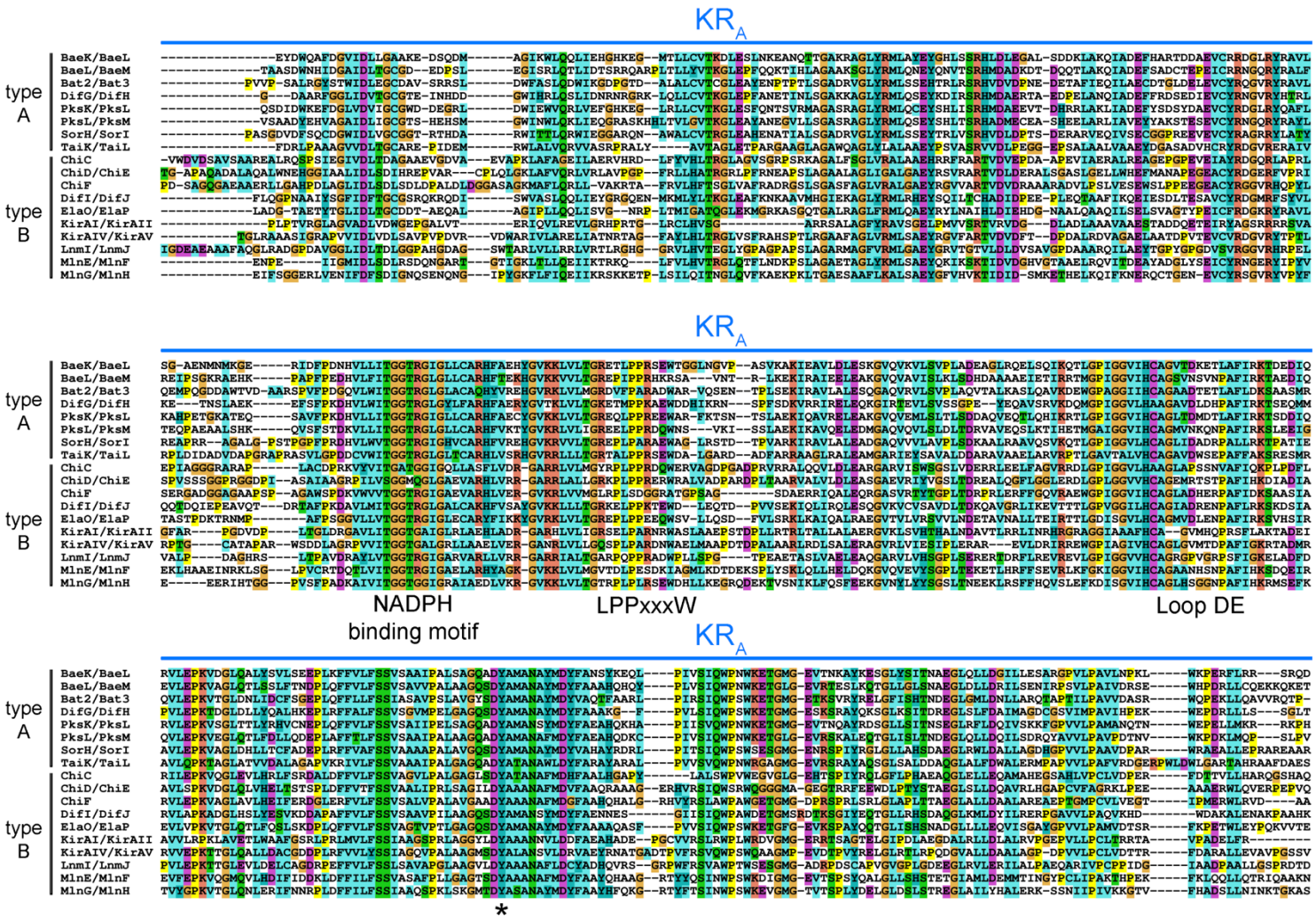
Figure 4-1 continued;

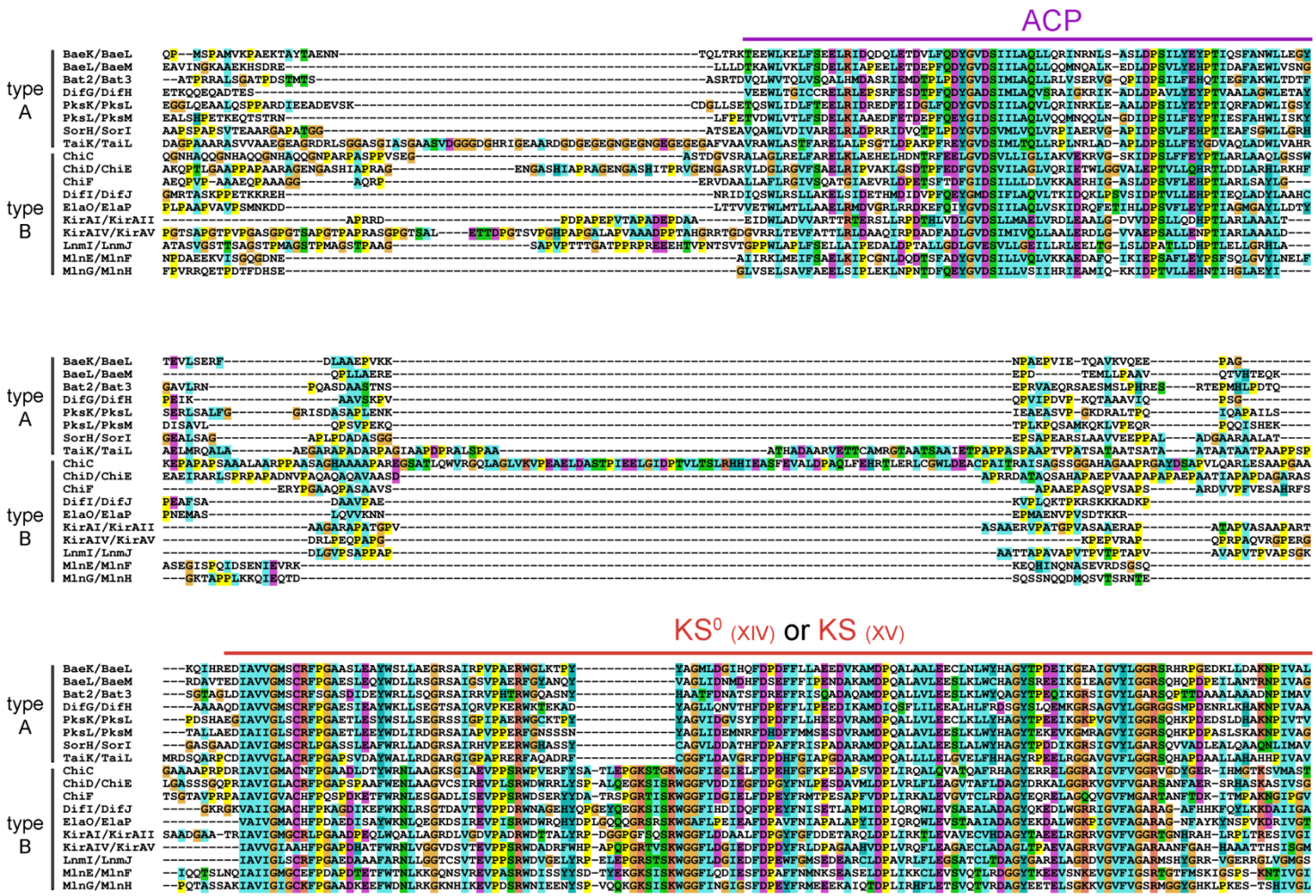
(B) Type B dehydrating bimodules are typically split into polypeptides containing KS + KR_A + ACP + KS and DH + ACP + KR_B. The KS that follows the first module belongs to clade XV, and a KS that follows the second module belongs to clade IX. DifDH10 is the DH within the second module of the type B dehydrating bimodule formed by the polypeptides DifI/DifJ from the difficidin assembly line of B. amyloliquefaciens FZB42. Asterisks show where the unsaturated moieties formed by the illustrated dehydrating bimodules appear in the fully synthesized natural product.

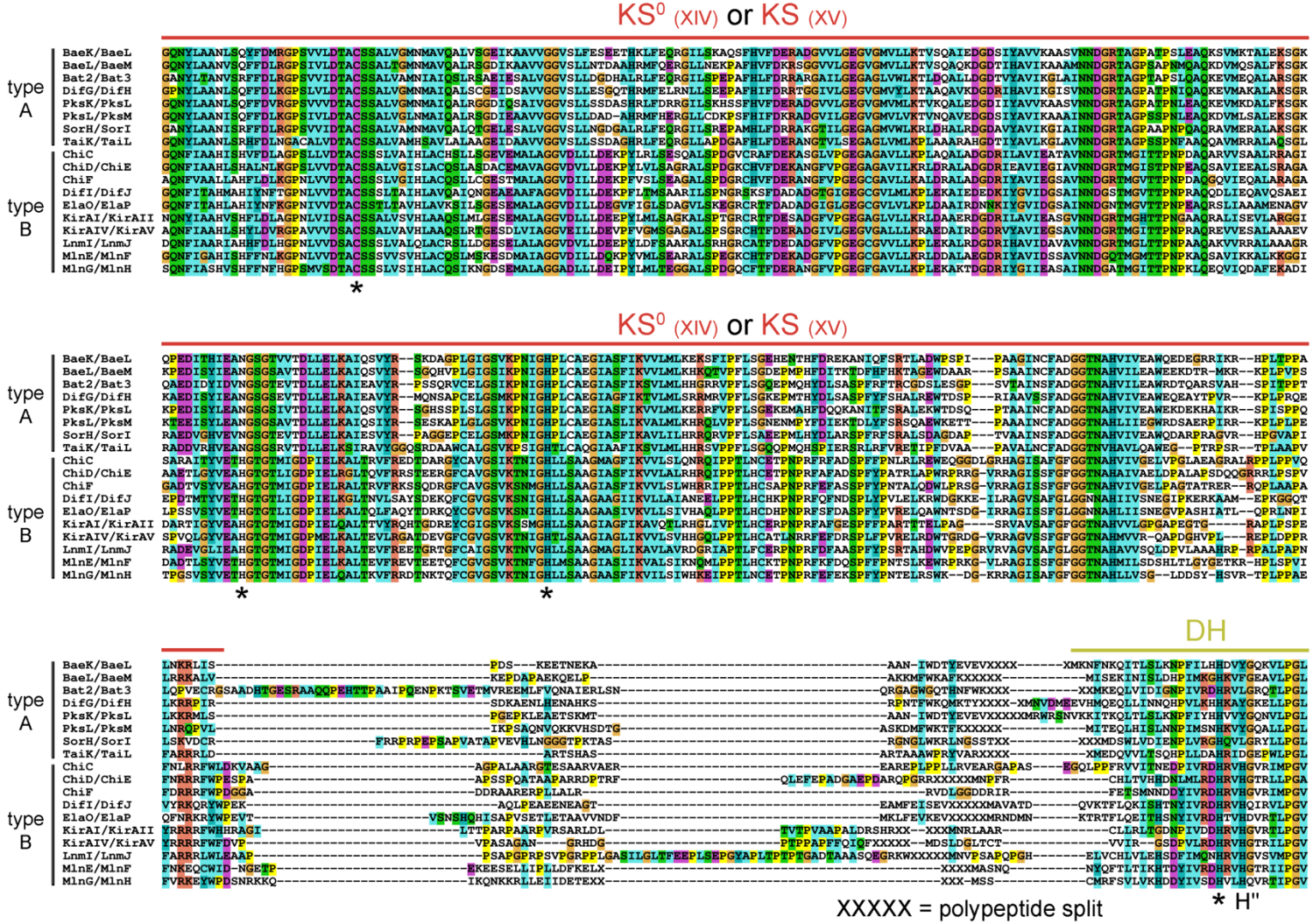
A)

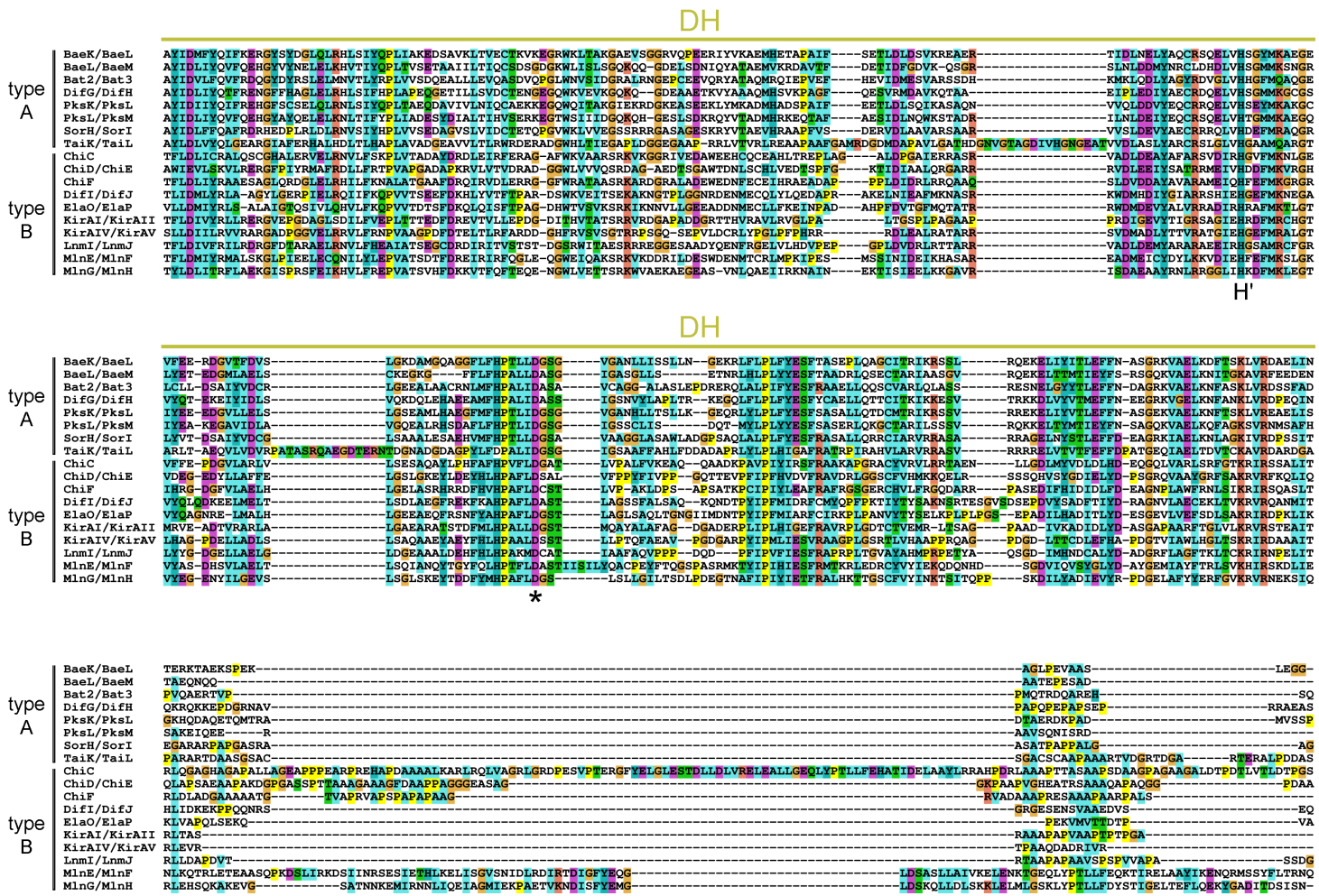

KS
KS
KS
type A
type B
type A
type B
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAlV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH

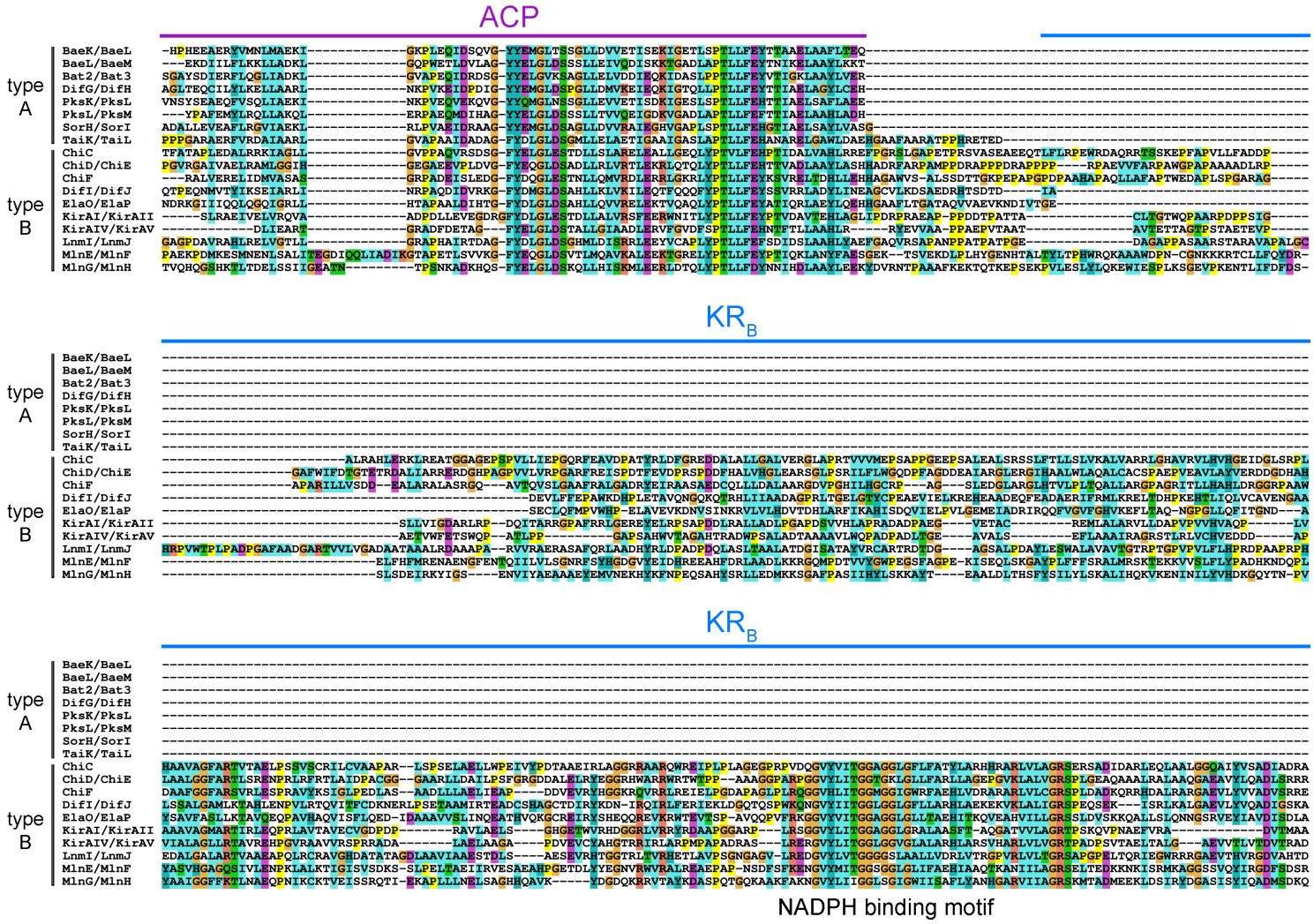

type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
flanking subdomain
(LINKS)
KR$_A$

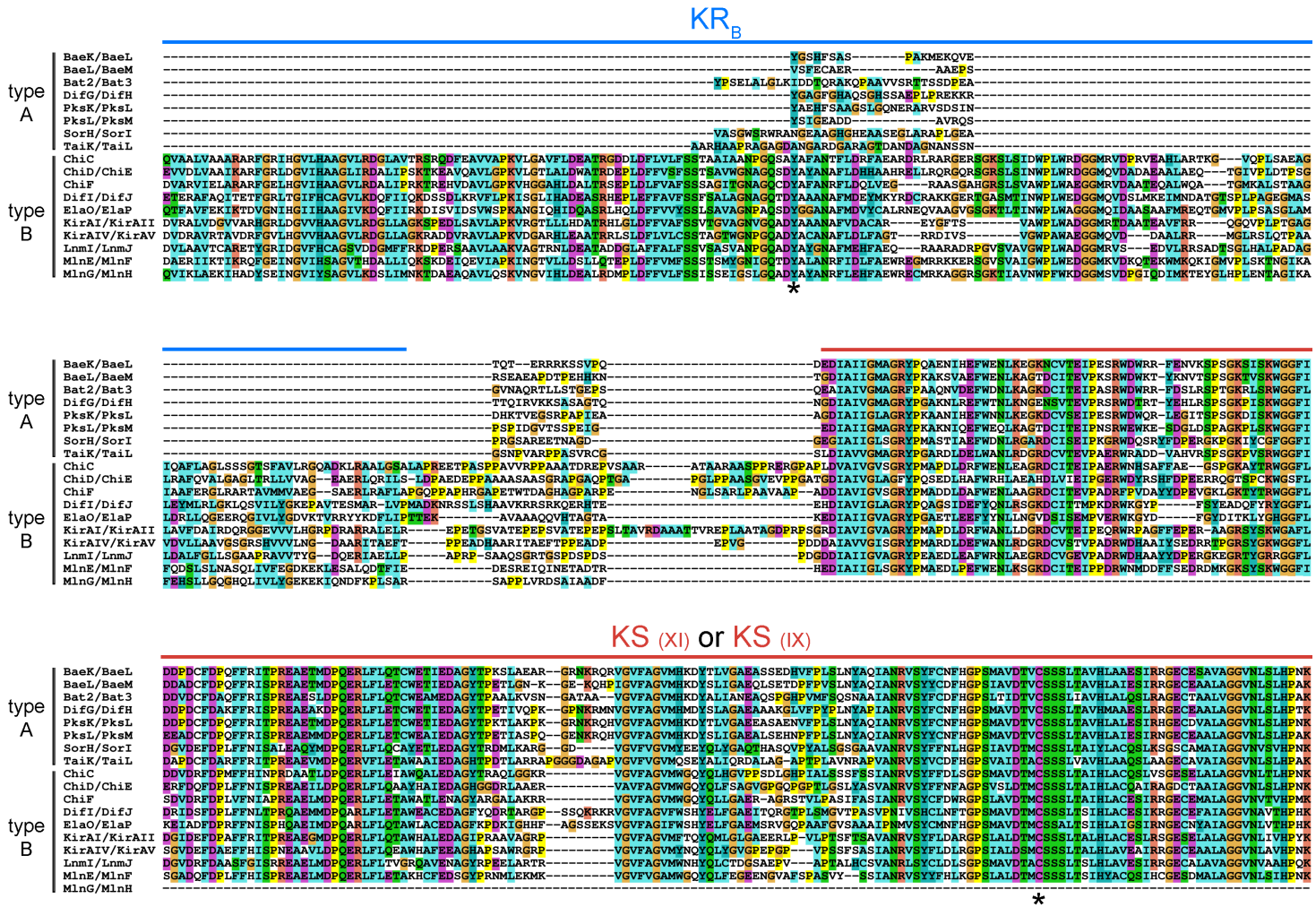

KR$_A$
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmJ/LnmJ
MlnE/MlnF
MlnG/MlnH
KR$_A$
NADPH binding motif
LPPxxxW
Loop DE
KR$_A$
*

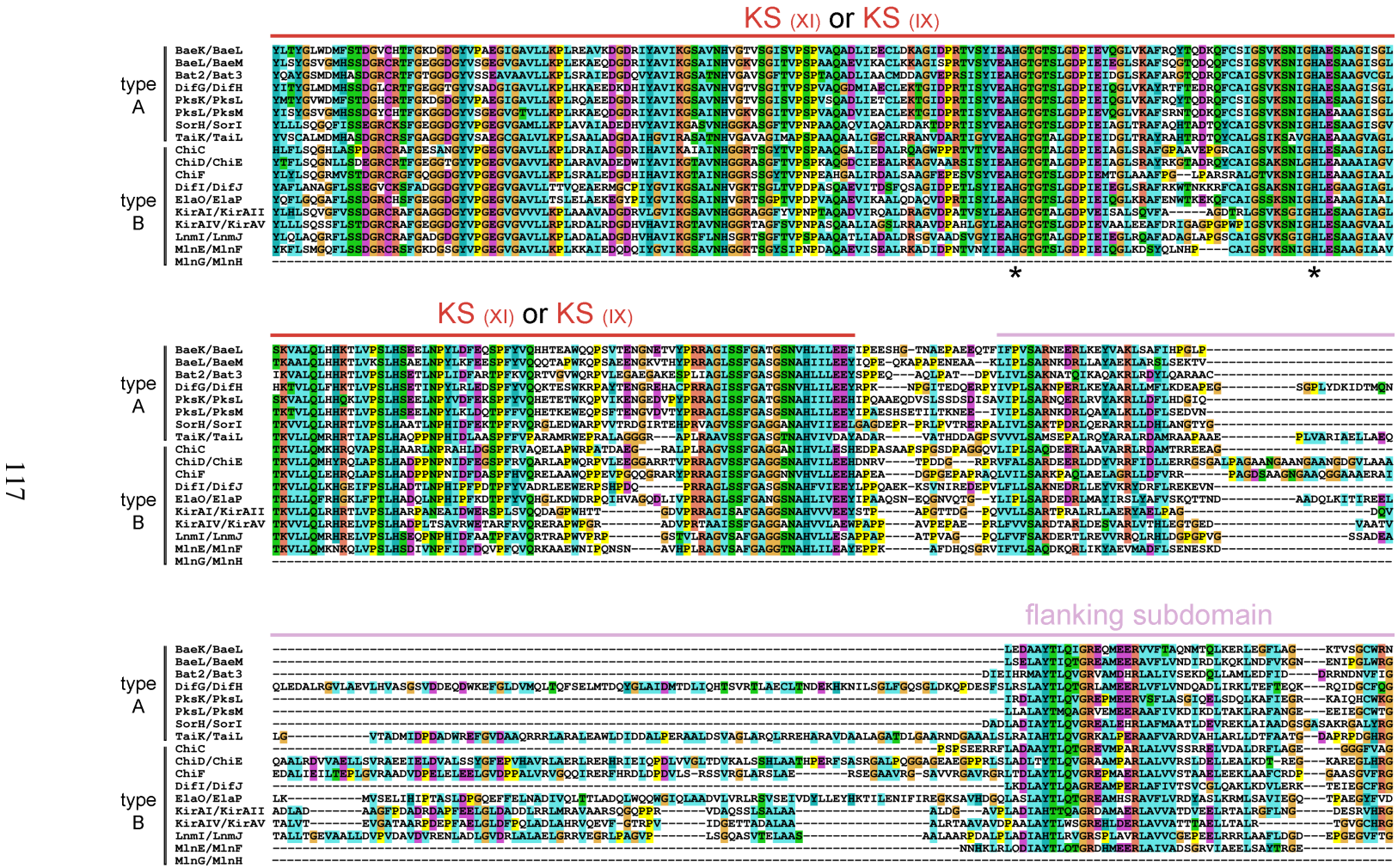

ACP
type A
type B
KS⁰ (XIV) or KS (XV)

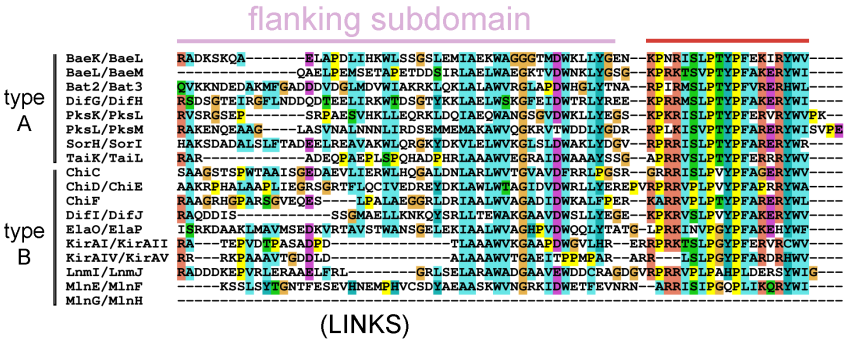

KS⁰ (XIV) or KS (XV)
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
KS⁰ (XIV) or KS (XV)
DH
XXXXX = polypeptide split
* H"


DH
type A
type B
H'
DH
type A
type B
*
type A
type B


ACP
KR$_B$
KR$_B$
type A
type B
type A
type B
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
NADPH binding motif


KR B
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
KS (XI) or KS (IX)


KS (XI) or KS (IX)
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
flanking subdomain


flanking subdomain
type A
type B
BaeK/BaeL
BaeL/BaeM
Bat2/Bat3
DifG/DifH
PksK/PksL
PksL/PksM
SorH/SorI
TaiK/TaiL
ChiC
ChiD/ChiE
ChiF
DifI/DifJ
ElaO/ElaP
KirAI/KirAII
KirAIV/KirAV
LnmI/LnmJ
MlnE/MlnF
MlnG/MlnH
(LINKS)

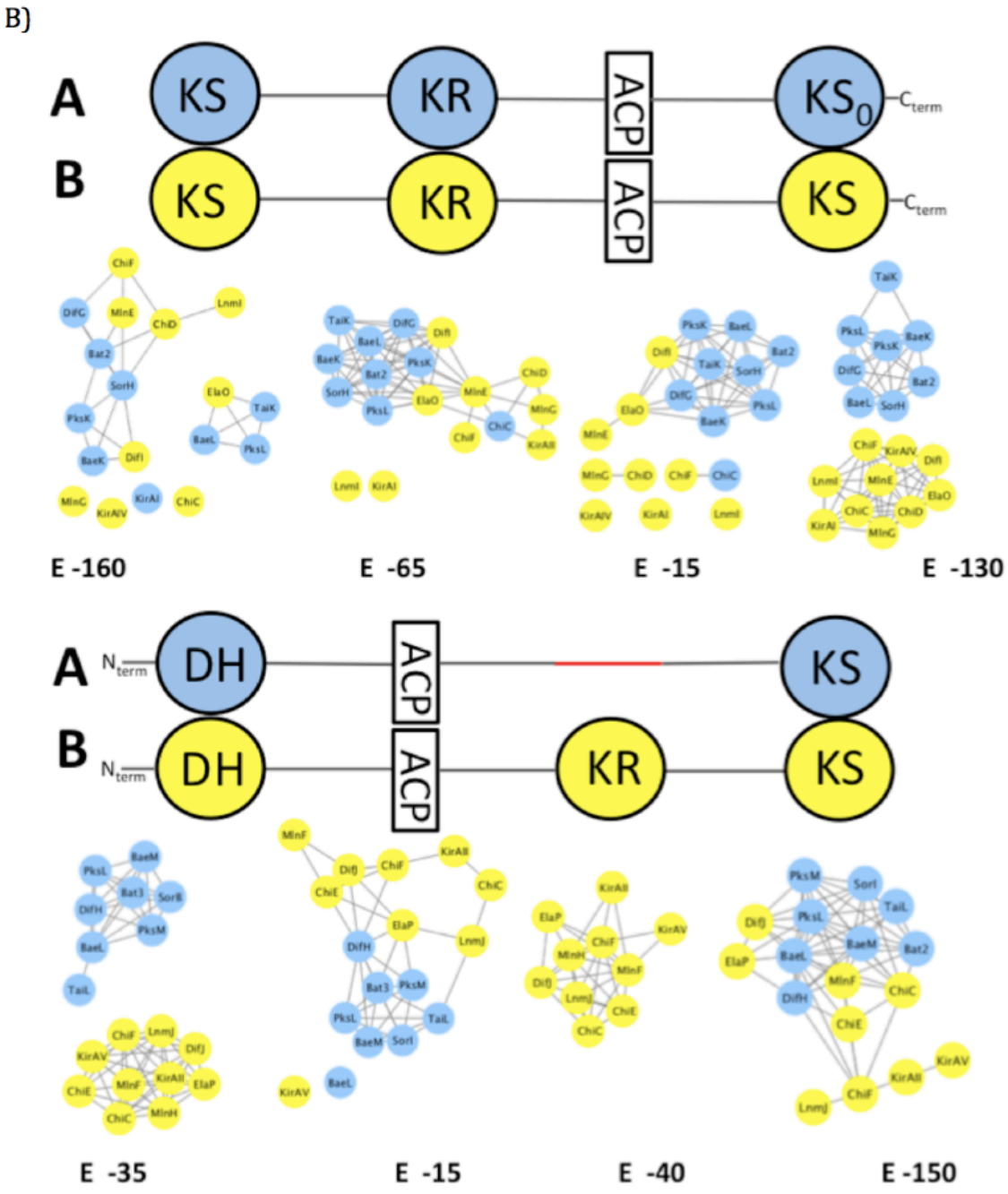


Figure 4-2: Comparing the Domains of Type A and Type B Bimodules.

Figure 4-2 continued;

(A) The sequences of dehydrating bimodules (and the ensuing KS, except for the MlnG/MlnH bimodule, which is followed by a TE) are aligned. Type A and B dehydrating bimodules have equivalent domain architecture until the KR_B of type B dehydrating bimodules. Domains are indicated with a line above the sequences colored as in Figure 1. Catalytic residues are indicated by asterisks, and motifs are denoted with the same text used in the article, except for the "LINKS" motif which is described elsewhere (Gay et al., 2016). BaeK/BaeL/BaeM, bacillaene, Bacillus amyloliquefaciens, CAG23957/CAG23958/CAG23959; Bat2/Bat3, batumin, Pseudomonas fluorescens, ADD82940/ADD82941; ChiC/ChiD/ChiE/ChiF, chivosazol, Sorangium cellulosum, AAY89050/AAY89051/AAY39052/AAY39053; DifG/DifH/DifI/DifJ, difficidin, Bacillus amyloliquefaciens, CAG23978/CAJ57408/CAJ57409/CAJ57410; ElaO/ElaP, elansolid, Chitinophaga sancti, AEC04361/AEC04362; KirAI/KirAII/KirAIV/KirAV, kirromycin, Streptomyces collinus, AGS73523/AGS73524/AGS73526/AGS73527; LnmI/LnmJ, leinamycin, Streptomyces atroolivaceus, AAN85522/AAN85523; MlnE/MlnF/MlnG/MlnH, macrolactin, Bacillus amyloliquefaciens, ABS73800/ABS73801/ABS73802/ABS73803; PksK/PksL/PksM, bacillaene, Bacillus subtilis, WP_014476859/WP_014476860/WP_014476861; SorH/SorI, sorangicin, Sorangium cellulosum, ADN68483/ADN68484; TaiK, TaiL, thailandamide, Burkholderia thailandensis, WP_009897624/WP_009897623.

Figure 4-2 continued;

(B) Sequence Similarity Networks show the relationships between the domains of type A and B dehydrating bimodules. The Enzyme Function Initiative–Enzyme Similarity Tool was used to identify similarity between the equivalent domains of type A(blue) and type B (yellow) dehydrating bimodules (Gerlt et al., 2015). Cut off alignment scores are indicated.

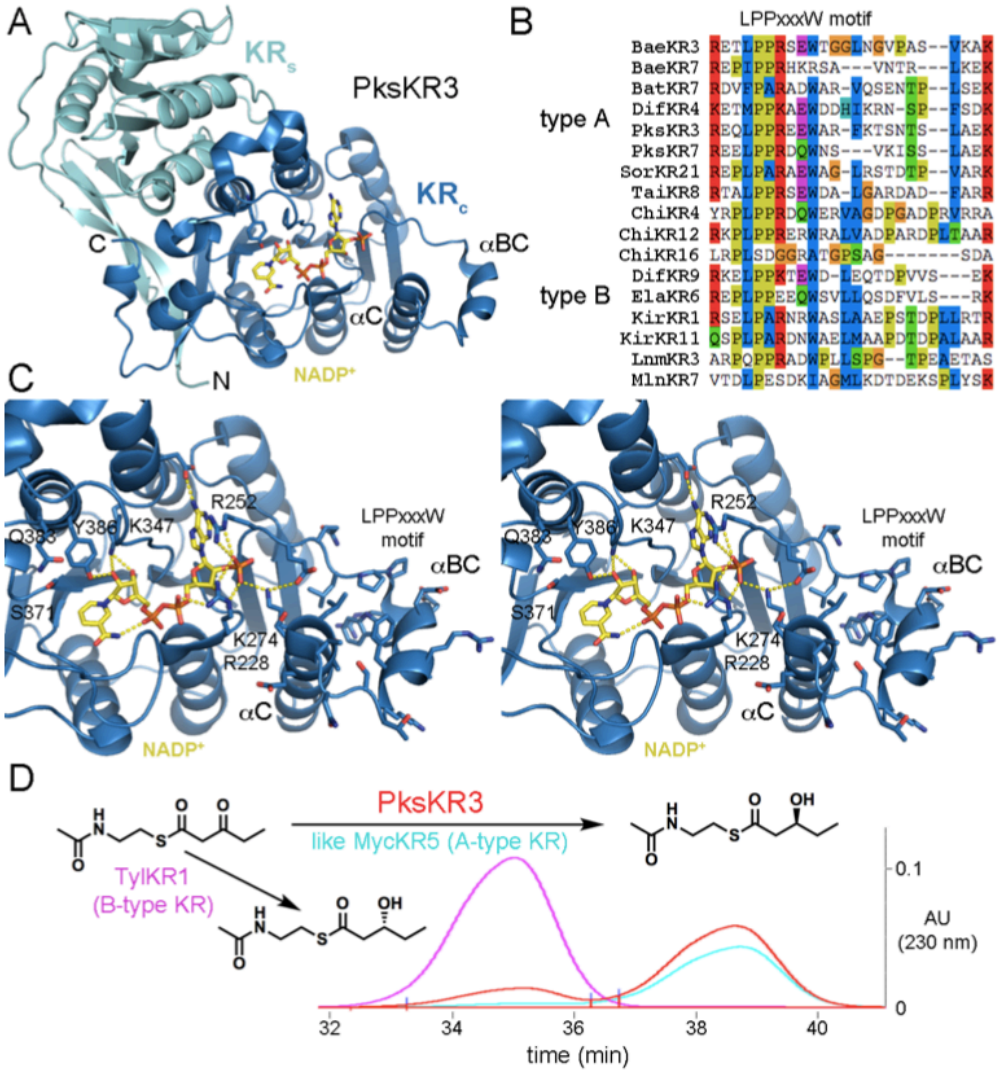


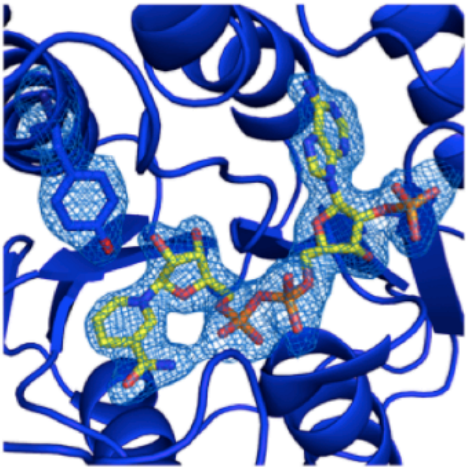
Figure 4-3: KR_A Structure and Function.

(A) The 1.98 A˚ resolution structure of PksKR3 bound to NADP+ reveals the unusual architecture of KRAs from dehydrating bimodules. The structural subdomain (KRs)

Figure 4-3 continued;

possesses an additional b strand, while the catalytic subdomain (KRc) possesses an additional a helix (aBC). (B) The LPPxxxW motif associated with aBC is partially conserved throughout the dehydrating bimodules. Type A: BaeKR3/ BaeKR7, bacillaene, Bacillus amyloliquefaciens, CAG23957/CAG23958; DifKR4, difficidin, Bacillus amyloliquefaciens, CAG23978; PksKR3/PksKR7, bacillaene, Bacillus subtilis, WP_014476859/ WP_014476860; SorKR21, sorangicin, Sorangium cellulosum, ADN68483; TaiKR8, thailandamide, Burkholderia thailandensis, WP_009897624. Type B: ChiKR4/ChiKR12/ChiKR16, chivosazol, Sorangium cellulosum, AAY89050/AAY89051/AAY89053; DifKR9, difficidin, Bacillus amyloliquefaciens, CAJ57409; ElaKR6, elansolid, Chitinophaga sancti, AEC04361; KirKR1/KirKR11, kirromycin, Streptomyces collinus, AGS73523/AGS73526; LnmKR3, leinamycin, Streptomyces atroolivaceus, AAN85522; MlnKR7, macrolactin, Bacillus amyloliquefaciens, ABS73800. (C) A stereodiagram of KRc of PksKR3 reveals several features near the NADPH-binding site. The catalytic tyrosine (Y386) and active site residues K347, S371, and Q383 are oriented similar to other structurally characterized KRs, while the interactions of R228 with the nicotinamide cofactor have not been observed in other KRs. The LPPxxxW motif, which includes the protruding aBC have also not been previously observed. (D) PksKR3 reduces the model substrate 3-ketopentanoyl-S-NAC as expected for the A-type mycolactone KR MycKR5 and not the B-type tylosin KR TylKR1 (Piasecki et al., 2011)

A)



B)

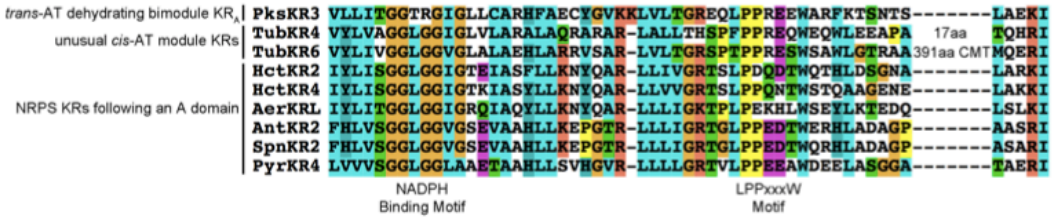


Figure 4-4: PksKR3 omit map and alignment of bimodule KRs.

(A) F_O-F_C omit map of $NADP^+$ cofactor and Y386 in PksKR3 active site.

(B) The unusual LPPxxxW motif of KR_As of *trans*-AT dehydrating bimodules, such as PksKR3, is also present in KRs from the tubulysin *cis*-AT PKS as well as KRs that follow A domains within NRPS assembly lines. A C- methyltransferase (CMT) inserts

Figure 4-4 continued;

into TubKR6 imFmediately following the LPPxxxW motif. The structure of a CMT fromthe curacin assembly line shows an LPPDFLL sequence at its N-terminus, suggesting that the LPPxxxW motif is the vestiges of a CMT (Skiba et al., 2016). AerKRL, aeroginosin, *Microcystis aeruginosa*, WP_069474084; AntKR2, antimycin, *Streptomyces sp. S4,* WP_010639240; HctKR2/HctKR4, *Lyngbya majuscula*, hectochlorin, AAY42397/AAY42398; PksKR3, bacillaene, *Bacillus subtilis*, WP_014476859; PyrKR4, pyridomycin, *Streptomyces pyridomyceticus*, WP_051829854: SpnKR2, splenocin, *Streptomyces sp. CNQ431*, WP_033947913; TubKR4/TubKR6, tubulysin, *Archangium disciforme*, CAF05649/CAF05651.

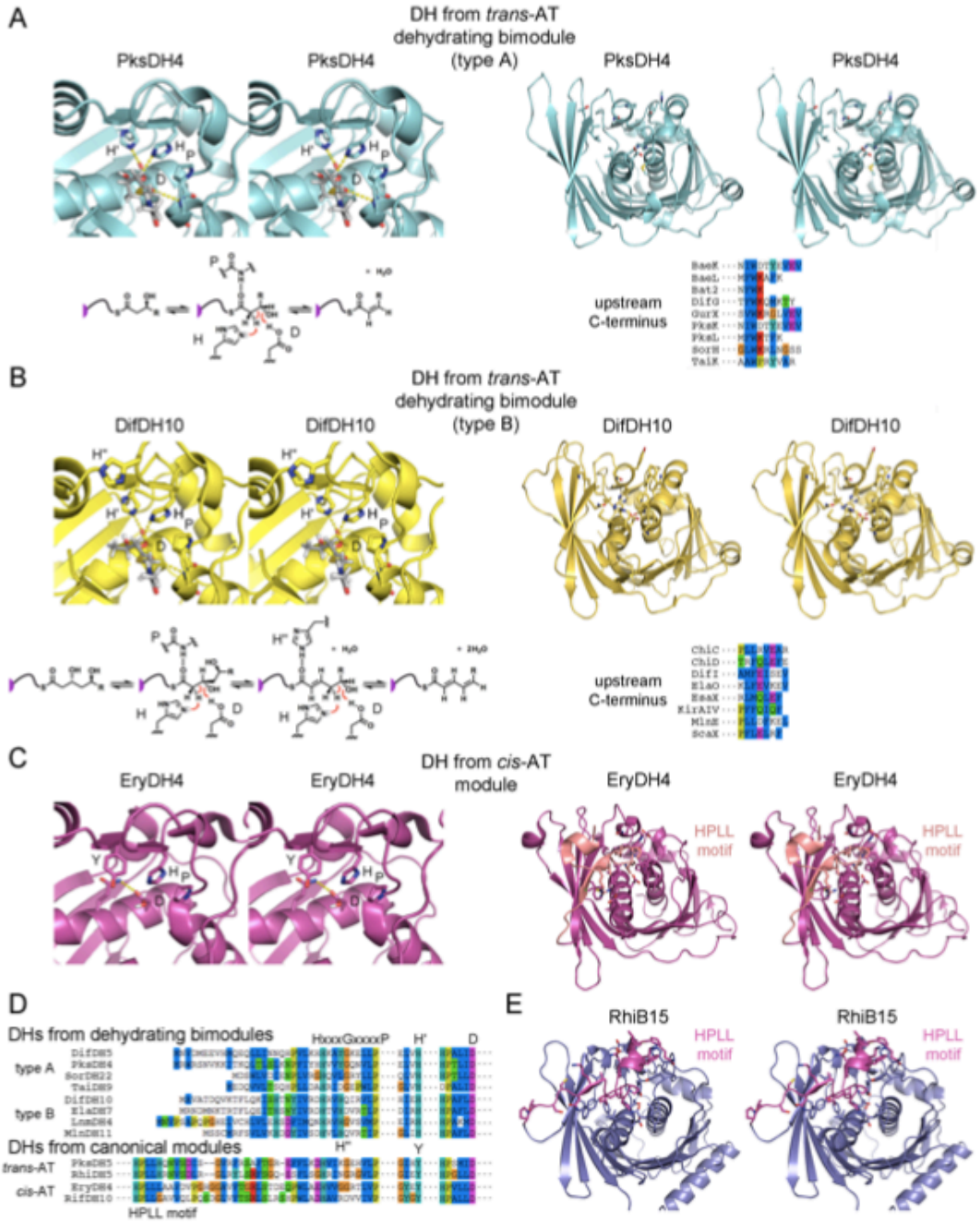


Figure 4-5: DH Structure and Function.

Figure 4-5 continued;

(A) PksDH4 from the B. subtilis 168 bacillaene assembly line (PDB: 5E1V) is representative of DHs from the second module of split type A dehydrating bimodules. Stereodiagrams show an L-b-hydroxyacyl substrate modeled in its active site as well as the putative docking site for the C terminus of the upstream polypeptide. The catalytic histidine and active site aspartate (H and D) may perform dehydration of L-b-hydroxyacyl substrates to a/b-cis double bonded products as proposed for DHs from cis-AT assembly lines (Keatinge-Clay, 2016). (B) DifDH10 from the difficidin assembly line is representative of DHs from the second module of split type B dehydrating bimodules. Stereodiagrams show a D-b-hydroxyacyl substrate modeled in its active site as well as the putative docking site for the C terminus of the upstream polypeptide (with the exception of ChiC, which is not split but contains a similar motif). The catalytic histidine and active site aspartate (H and D) may perform dehydration of a D-b-L-d-dihydroxyacyl substrate to yield an a,b-trans double bond as proposed for DHs from cis-AT assembly lines. The histidine that abstracted the proton from the L-a-hydrogen could again abstract a proton from the L-g-hydrogen as the aspartate helps donate a proton to the L-d-hydroxyl group, as shown in Route 1. A histidine conserved in DHs from type A dehydrating bimodules (H00) may be too distant to directly aid catalysis. The alternative Route 2 is based on how the DH in the final module of the ajudazol PKS is thought to mediate dehydration of a b-keto, d-hydroxy intermediate to yield a g/d-cis double bond. This relies on the presence of the b-keto group, which may positioned similar to a

Figure 4-5 continued;

thioester carbonyl during canonical dehydration, and thus proceed before the reduction and dehydration that generates the a/b-trans double bond. (C) EryDH4 from the erythromycin assembly line (PDB: 3EL6) is representative of DHs from cis-AT assembly lines. Stereodiagrams show its active site with a Y in the same position as the signature H0 of DHs from dehydrating bimodules as well as the interactions of the N-terminal HPLL motif with the rest of the DH fold. The WPP motif of cis-AT DHs is replaced by a short helix in trans-AT DHs. (D) Sequences of DHs from split type A and B dehydrating bimodules, as well as canonical DHs from trans-AT and cis-AT assembly lines, are aligned. DHs from split dehydrating bimodules do not possess the HPLL motif of canonical DHs; the upstream C termini shown in (A and B) may dock to the same hydrophobic pocket as the HPLL motif (see Figure 4). The locations of the proton-abstracting histidine (H), the signature histidine (H0) that replaces the tyrosine (Y) of canonical DHs, the histidine of type B dehydrating bimodules (H00), and the active site aspartate (D) are indicated. (E) A B domain from the rhizoxin assembly line (RhiB15, PDB: 4KC5) reveals how an HPLL motif completes the fold of that DH-like domain. The polypeptide immediately upstream of this motif also interacts with an N-terminal KS to rigidly connect the dimeric B domain with the dimeric KS. BaeK/BaeL, bacillaene, Bacillus amyloliquefaciens, CAG23957/CAG23958; Bat2, batumin, Pseudomonas fluorescens, ADD82940; ChiC/ChiD, chivosazol, Sorangium cellulosum, AAY89050/AAY89051; DifG/DifI/DifDH5/DifDH10, difficidin, Bacillus

Figure 4-5 continued;

amyloliquefaciens, CAG23978/CAJ57409/CAJ57408/CAJ57410; ElaO/ElaDH7, elansolid, Chitinophaga sancti, AEC04361/AEC04362; EryDH4, erythromycin, Saccharopolyspora erythraea, AAV51821; EsaX, uncharacterized, Enhygromyxa salina, KIG11581; GurX, uncharacterized, Geobacter uraniireducens, ABQ27263; KirAIV, kirromycin, Streptomyces collinus, AGS73526; LnmDH4, leinamycin, Streptomyces atroolivaceus, AAN85523; MlnE/MlnDH11, macrolactin, Bacillus amyloliquefaciens, ABS73800/ABS73803; PksK/PksL/PksDH4/PksDH5, bacillaene, Bacillus subtilis, WP_014476859/WP_014476860; RhiDH5, rhizoxin, Burkholderia rhizoxinica, CAL69889; RifDH10, rifamycin, Amycolatopsis mediterranei, WP_013222551; ScaX, uncharacterized, Streptomyces cattelya, WP_014144300; SorH/SorDH22, sorangicin, Sorangium cellulosum, ADN68483/ADN68484; TaiK/TaiDH9, thailandamide, Burkholderia thailandensis, WP_009897624/WP_009897623.

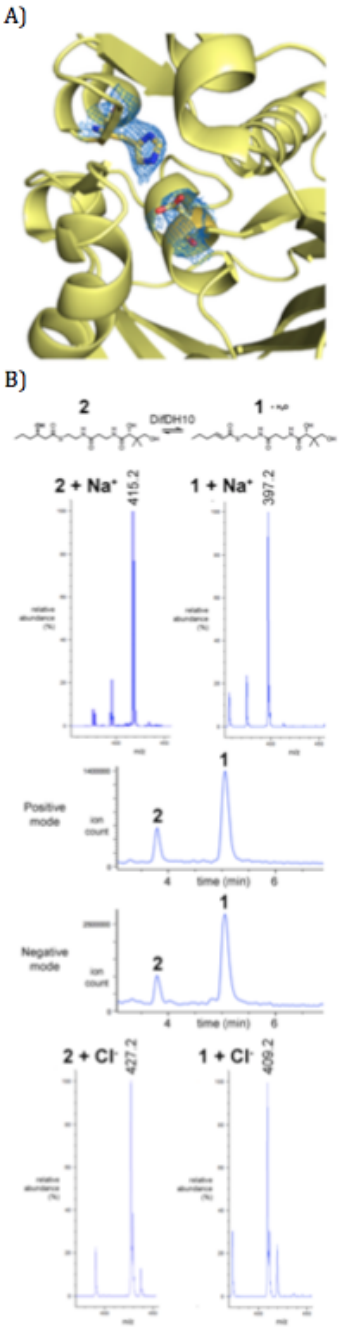


Figure 4-6: DifDH10 omit map and dehydration assay.

Figure 4-6 continued;

(A) F_o- F_c omit map of active site residues H25 and D198 in the active site DifDH10.

(B) DifDH10 was incubated with (*E*)-hex-3-enyl-*S*-pantetheine (1) for 4 h and analyzed as described. The LC/MS data, both in positive and negative modes, shows DifDH10 catalyzes the hydration of 1 to 2.

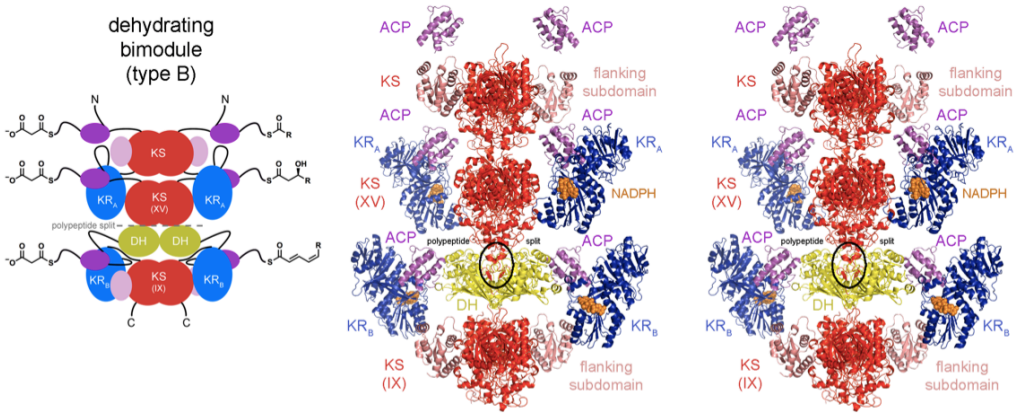


Figure 4-7: Model of a Dehydrating Bomodule.

A stereodiagram shows a model for a split type B dehydrating module generated entirely from structurally elucidated trans-AT assembly line domains. The structure of PksKS2 from the B. subtilis 168 bacillaene assembly line (PDB: 4NA1) was used for the first KS as well as for the KS following the bimodule, while the structure of the Rhi(KS + B)15 didomain from the rhizoxin assembly line (PDB: 4KC5, without the flanking subdomain) was utilized for the KS of the second module (Gay et al., 2014a; Bretschneider et al., 2013). The structure of the DH-like B domain RhiB15 (PDB: 4KC5) serves as a model for DH (Bretschneider et al., 2013). The structure of PksKR3 reported here represents KRA, and the structure of MlnKR6 from the macrolactin assembly line (PDB: 5D2E) represents KRB (Zeng et al., 2016). The structure of MupACP7a from the mupirocin assembly line (PDB: 2L22) provides templates for the ACPs (Haines et al., 2013). All

Figure 4-7 continued;

domains except for the C-terminal KS and the N-terminal DH of split dehydrating bimodules are connected by flexible linkers as indicated by the schematic on the left. The location of the polypeptide split is circled in black. Interactions between the C terminus of the first polypeptide and a hydrophobic pocket in the N-terminal DH of the second polypeptide are hypothesized to facilitate docking.

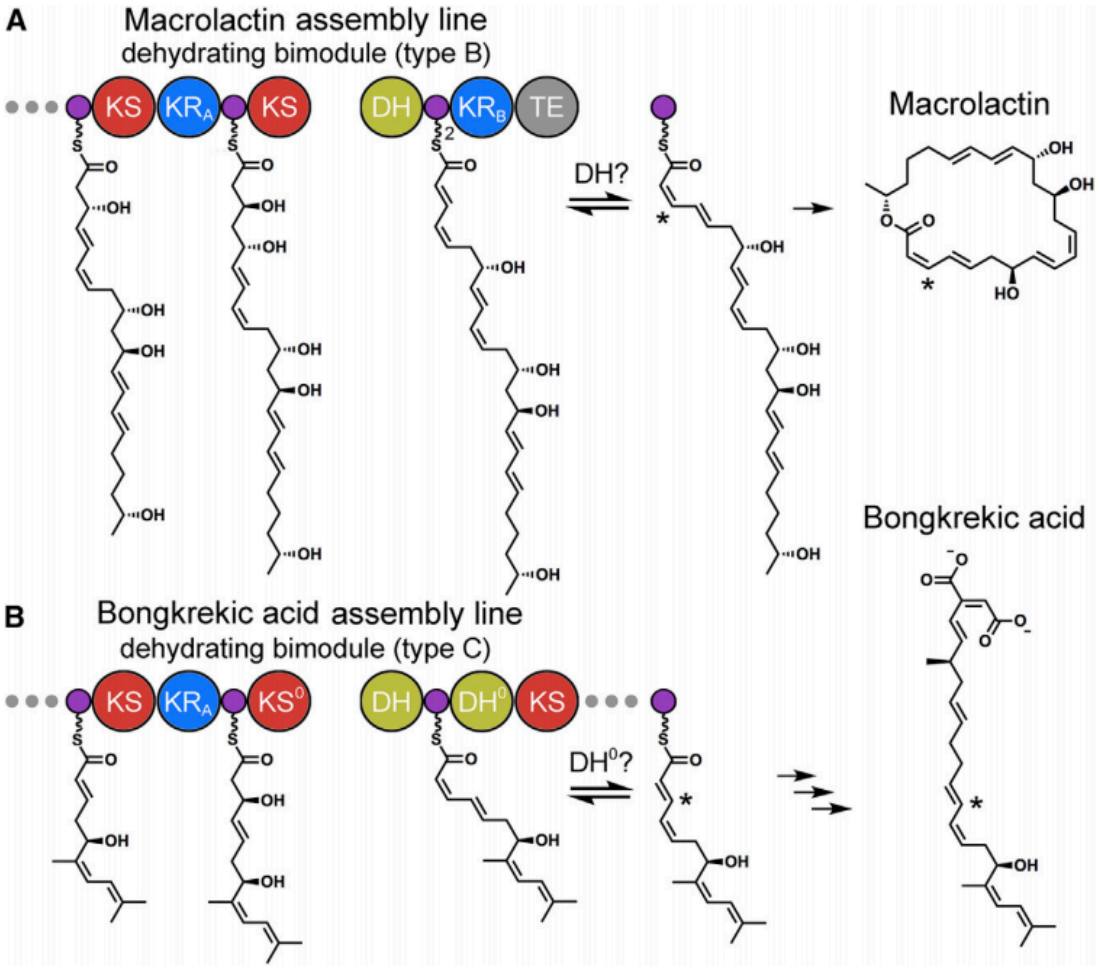


Figure 4-8: Mysterious Isomerization.

Accurate prediction of polyketide intermediates generated by dehydrating bimodules is thwarted by a mysterious isomerization activity. (A) This phenomenon is best illustrated in the macrolactin assembly line, where an offloading thioesterase (TE) immediately follows a type B dehydrating bimodule. The a/b-trans-g/d-cis diene anticipated from the

Figure 4-8 continued;

type B dehydrating bimodule is apparently converted to a a/b-cis, g/d-trans diene prior to macrocyclization. Perhaps DH catalyzes this isomerization, and TE selects the isomerized intermediate for macrocyclization.

(B) The interconversion of an a/b-cis, g/d-trans with an a/b-trans, g/d-cis diene is evidently mediated by a DH-like isomerase (labeled as DH0) in type C dehydrating bimodules. The KS following this isomerase may select the isomerized product for further elongation into a polyketide product such as bongkrekic acid. Asterisks show where the unsaturated moieties formed by the illustrated dehydrating bimodules appear in the fully synthesized natural product.

Table 1. Crystallographic Parameters

Data collection	**PksKR3**	**DifDH10**
Wavelength (Å)	1.0332	1.0332
Space group	$P2_12_12_1$	P1
Cell dimensions, *a, b, c* (Å)	46.1, 84.9, 134.3	41.4, 74.0, 94.2
Resolution (Å)	50.00 – 1.98	39.79 – 2.22
R_{merge}	0.082 (0.681)	0.104 (0.348)
$I/\sigma(I)$	21.8 (2.3)	6.5 (1.5)
No. of reflections	34,894	49,121
Completeness (%)	97.1 (87.2)	89.5 (61.4)
Redundancy	6.4 (5.6)	1.8 (1.4)
Refinement		
Resolution (Å)	71.74 – 1.98	39.79-2.22
No. of reflections	34,894	49,098
R_{work}/R_{free}	0.214/0.241	0.217/0.249
No. of atoms	3,811	9,816
Protein	3,560	9,353
Water	155	463
Average *B* factors (Å^2)		
Protein	25	29
Water	36.37	31.62
RMS deviations		
Bond lengths (Å)	0.01	0.032
Bond angles (°)	0.1602	2.478
Ramachandran Statistics (%)		
Preferred Regions	97.53(435)	96.50
Allowed Regions	2.47(11)	3.50
Outliers	0	0

Table 4-1: PksKR3 and DifDH10 crystallographic parameters.

Chapter 5: Structure and Activity of an Embedded Pyran Synthase Domain from a *trans*-Acyltransferase Polyketide Sythase

ABSTRACT

trans-Acyltransferase polyketide synthases often utilize obscure enzymatic domains not generally observed in the biosynthetic pathways of their *cis*-acyltransferase counterparts. Within this repertoire of largely unexplored biosynthetic chemistry lies a novel class of pyran synthase that catalyzes the formation of 5 and 6 membered cyclic ethers within a variety of polyketide systems. The 1.55 Å crystal structure of an embedded pyran synthase domain from the enigmatic sorangicin pathway highlights the similarity of this enzyme to the ubiquitous dehydratase domain and reveals new insights into the mechanism of pyran ring formation. Mutagenesis of active site residues corroborates the structural implications that together point toward a catalytic mechanism wherein the dehydratase catalytic dyad has been altered to create a novel active site centered around a conserved histidine as the sole residue directly involved in pyran ring formation.

INTRODUCTION

Polyketide natural products offer rich examples of the molecular complexity assembled by Nature from simple cellular metabolites. The sophistication of polyketides is realized through the action of the enormous modular mega-enzymes polyketide synthases (PKSs) (Keatinge-Clay, 2012). In prokaryotes, modular Type I PKS systems act as non-iterative assembly lines comprised of repeated multi-domain modules that

each perform a two-carbon extension and subsequent functionalization of a growing polyketide intermediate that ultimately produces a molecular scaffold that possessing exquisite bioactivity. The essential domains that define a module are the ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) necessary for selection, transfer and condensation of extender units onto nascent polyketides. A module also generally contains a sub-set of optional processing domains commonly including a ketoreductase (KR) for β-keto reduction, dehydratase (DH) for water elimination yielding an α/β–olefin and enoyl-reductase (ER) for reduction of the DH derived α/β–olefin (Keatinge-Clay, 2012). Several less commonly employed domains, such as the methyltransferase (MT) are used to further expand polyketide diversity.

Type I PKS systems are further eponymously classified as either *cis* or *trans*-acyltransferase based on whether AT domains are encoded in-line or as discrete *trans*-acting components that dock to the module(s) they supply extender units to. For the *trans* systems in particular, much remains to be learned about their underlying chemical logic as well as the scope of chemistry they employ (Piel, 2009; Piel, 2016). In addition to the traditional embedded domains they share with their *cis*-AT counterparts (KR, DH, ER, MT), *trans*-AT synthases often house domains of novel and poorly understood function involved in varied processes including O-methylation, β-keto methylation, double bond migration and pyran/furan ring formation. In aggregate, these additional domains allow for greater chemical diversity within the *trans*-AT realm and contribute to the variety and bio-activity seen within *trans*-AT polyketides.

Of the novel enzymes unique to *trans*-AT biosynthetic gene clusters are pyran synthase (PS) domains. Relative to *cis*-AT derived polyketides, *trans*-AT encoded molecules more commonly contain pyran and furan rings, due in part to their utilization of embedded PS domains dedicated to the installation of five- and six-membered ether rings onto polyketide backbones. Though several examples exist of pyran containing *cis*-AT polyketides such as ambruticin and salinomycin, they rely on machinery distinct to the PS domains harbored within *trans*-AT systems. In the case of *cis*-AT synthases, two separate routes to pyran rings are known. In ambruticin biosynthesis, a bifunctional DH domain carries out a sequential dehydration and cyclization that leads to formation of a tetrahydropyran ring (Berkhan et al., 2014). Remarkably, sequence level comparisons between AmbDH3 and typical mono-functional DH domains show no clear alterations to the active site architecture of AmbDH3 in comparison with the canonical DH His/Asp catalytic dyad retained in the expected orientation. Alternatively, structural investigation of the post-PKS processing enzyme from the salinomycin pathway, SalBiii, has revealed it to be evolutionarily distinct from the dehydratase domain with high homology to an epoxide hydrolase/cyclase like family of enzymes (Luhavaya et al., 2015). Based on this homology SalBiii is currently hypothesized to employ a catalytic dyad consisting of two key aspartates that facilitate ether ring formation.

Many *trans*-AT pathways use PS domains during polyketide biosynthesis to generate end products often possessing anti-proliferative or anti-microbial properties. Pathways utilizing PS domains include the pederin, psymberin, bryostatin, misakinolide, spliceostatin, ooocydin, thailanstatin and sorangicin systems with the thailanstatin and

sorangicin synthases encoding 2 and 3 PS domains respectively (Figure 5-1A). Of these domains, all naturally catalyze formation of a six-membered ring with the exception of the oocydin PS that produces a five-membered furan using a hydroxyl seemingly installed by a *trans*-acting monooxygenase (Matilla et al., 2012). The relative paucity of furans likely at least in part stems from their requirement for non-KR derived hydroxyl nucleophiles. Notably, a putative PS domain within the thailanstatin NRPS-PKS hybrid pathway uses a threonine side-chain hydroxyl as a nucleophile in forming a six-membered ring (Eustaquio et al., 2014; Zhang et al., 2011). It is presently unclear the extent to which PS domains display substrate specificity as well as if any distinction exists between five and six member ring forming PS domains.

Previous work on the excised pyran synthase (PS) domains from the pederin and misakinolide biosynthetic gene clusters (BSG) demonstrated cyclization activity towards α/β– unsaturated N-acetylcysteamine (SNAC) substrates indicating that PSs act immediately after the DH domain during chain elongation (Poplau et al., 2013, Helfrich et al., 2015). NMR analysis of the purified products showed that with regard to the newly formed C-O bond, the pederin PS closes the ring robustly and stereoselectively *in vitro* making PS domains potentially useful as tools for the chemoenzymatic synthesis of cyclic ether moieties. Although the evidence available suggests that PS domains close rings stereoselectively, inspection of the stereochemistry of the pyran rings within sorangicin A indicates that the 3 PS domains encoded by the BSG do not all add to the same side of their substrate's site of unsaturation. Though no structural information on

embedded PSs exists, sequence alignment reveals high similarity to the canonical DH domain, ubiquitous among both *cis*- and *trans*-PKSs, with subtle deviations occurring at critical positions within the DH active site.

The sorangicin pathway of *Sorangium cellulosum* is a large and poorly understood *trans*-AT pathway that generates antibiotics possessing potent broad-spectrum activity (Irschik et al., 1987). The sorangicins are unique for their inclusion of three PS derived tetrahyropyran rings within their macrolide scaffolds (Figure 5-1A) (Irschik et. Al., 2010). Bioinformatic analysis of the sorangicin BSG indicates that each of the three pyran rings is installed by a dedicated in-line pyran synthase domain. Though each PS domain is found within a slightly varied modular context (DH KS^0PS ACP; KS DH PS KR ACP; KS DH PS --- KR ACP) they all are preceded by an active DH to provide them with the α/β–unsaturated substrate for the oxa-conjugate addition they catalyze (Figure 5-1B) (Zhang et al., 2017; Keatinge-Clay, 2017). Belying the considerable DH homology exhibited by PS domains, the Michael addition reaction they catalyze is mechanically distinct from the classic α/β-dehydration reaction that typifies the DH. Thus, the pyran synthase active site is hypothesized to have diverged in subtle ways in order to facilitate the intramolecular addition of a hydroxyl group across an α/β-olefin instead of catalyzing the removal of water from across a double bond. To shed light on the structural and mechanistic details of this as-of-yet un-elucidated enzyme, we report the first in class structure of a pyran synthase domain, SorPS9, from the ninth module of the sorangicin assembly line, alongside functional, mutational and

bioinformatic analysis of the SorPS9 active site.

RESULTS AND DISCUSSION

The identification of the boundaries of SorPS9 was accomplished via a multiple sequence alignment of several pyran synthase sequences alongside multiple structurally characterized dehydratase domains. This yielded the amino acid fragment consisting of residues 1360-1641 of the SorB polypeptide whose corresponding gene region was PCR amplified from purified *S. cellulosum* SoCe12 genomic DNA and cloned into the pET28b expression vector for protein purification & crystal screening trials. Diffraction of the optimized crystals yielded a 1.55 Å structure solved via molecular replacement using the CurDH8 structure (PDB: 3KG9) as a search model (Table 5-S1) (Akey, Structure 2010).

Similar to other recently solved *trans*-AT DH-like domains SorPS9 crystallized as a monomer with its overall structure closely resembling the canonical DH architecture characterized by the double-hotdog fold topology (Figure 5-2A) (Akey et al., 2010; Gay et al., 2014 Zeng et al., 2016; Gay et al., 2013; Keatinge-Clay, 2008). Compared with recently solved *trans*-AT DH domains, MlnDH2 and RzxBDH (PDB: 5HST and 5IL6) SorPS9 displays RMSD Cα values of 1.4 Å and 1.6 Å, respectively, indicating that PS domains possess a high overall structural homology to *trans*-AT DH domains (Figure 5-3). The active site of SorPS9 however, contains several divergent features not observed in any dehydratase structure solved to date.

Despite the irregularity of module nine of the sorangicin synthase, SorPS9 appears representative of other PS domains with its novel features shared by other known

PS domains (Figures 5-1B, 5-1C & 5-4) (Eustáquio et al., 2014; Pöplau et al., 2013; Helfrich at al., 2015; Xavier et al., 2014). The active site of SorPS9 is related to the prototypical DH active site, yet it contains distinct features in several critical aspects. Most notably, the highly-conserved His-Asp catalytic dyad used by the DH to facilitate water elimination is disrupted at the aspartate position in SorPS9 as well as other PS domains. In the DH mechanism, the aspartate is hypothesized to act as a proton donor to generate the leaving water, but in SorPS9, an asparagine, N186, is found in its place (Figure 5-2). Substitution at this position appears highly conserved for all PS domains but occasionally, as in the PSs harbored within the 16th module of the sorangicin synthase and the 9th module of the bryostatin pathway, a histidine is found at this position instead (Figure 5-4). Additionally, putatively assigned PS domains from the related thailanstatin and spliceostatin systems that lack other common PS motifs possess either a serine or phenylalanine at this position instead (Figure 5-5)(Eustáquio et al., 2014). Though the Cα of N186 superposes well with the aspartate backbone of the DH, the amide of the carbonyl side chain of SorPS9 is shifted closer to the catalytic histidine H33 (Figure 5-4B). This reorientation could allow hydrogen bonding with the ζ-hydroxyl group of the substrate or possibly with the imidazole of H33. H33 and N186 reside in a large hydrophobic cleft devoid of other ionizable residues. The hydrophobicity of this environment may play a role in the exclusion of water or in decreasing the pKa of the catalytic histidine. The scarcity of surrounding polar functional groups also limits the number of residues capable of participating in catalysis implying a mechanism primarily dependent upon H33 and N186.

Unique to PS domains is a four residue deletion within the HX_8P motif conserved among all DH domains (Keatinge-Clay, 2008). The final proline of the motif is highly conserved among both PS and DH domains and is thought to stack against the catalytic histidine to align the imidazole for abstraction of the substrate α-hydrogen (Gay et al, 2014). The truncation to this motif occurs in an external loop immediately after the active site histidine but does not significantly alter the side-chain positions of H33 or P46. Despite this, the truncation does prevent the formation of 2 β-strands observed in other DH structures that form the back side of the PS active site (Figures 5-2, 5-4 & 5-3). The removal of this secondary structure may confer increased flexibility to the active site cavity to facilitate substrate rearrangement during cyclization. This deletion appears unique to and conserved across PS domains and is the only significant alteration to secondary structure that sets SorPS9 apart from DH domains.

A structural alignment of SorPS9 with the recently structurally characterized *trans*-AT DH from *Pseudomonas fluorescens* (PDB:5IL6) and the enoyl-isomerase (EI), PksEI16 (PDB: 4U3V) from *Bacillus subtilis* reveals that SorPS9 possesses elements in common with both the DH and the EI (Gay et al., 2014; Kusebauch et al., 2010). SorPS9's active site aspartate to asparagine substitution is also present in the PksEI16 active site but its role in EI double bond migration is currently unclear. As the Asp to Asn mutation in PksEI16 superposes more closely with the Asp of the DH, N186 may have assumed a novel role during pyran ring formation (Fig 5-2 and 5-3). The imidazole side chains of the PS and DH catalytic histidines superpose with each other but in PksEI16 the catalytic histidine is shifted by ~1.7 Å which is postulated to be due to the absence of the

aforementioned conserved proline at the end of the HX_8P motif. The absence of this proline seems to confer greater mobility to the catalytic histidine of PksEI16, facilitating its role as proton shuttle between the γ–carbon and α–carbon. This proline is retained within the PS truncated HX_4P motif. restricting the movement of H33 within the active site. A conserved HPLL motif is generally found at the N-terminus of all three domain types yet it is only the EI that is missing the β-strand that follows the HPLL motif.

Bioinformatic inspection of KRs preceding PS domains suggests that all PS domains act on *trans*-α/β unsaturated intermediates. The configuration of the double bond acted upon by SorPS9 can be inferred from of the stereochemistry of the hydroxyl set by the upstream KR as hydroxyl orientation determines whether dehydration generates a *cis* or *trans* α/β–olefin. Sequence analysis of KRs upstream of PS domains reveals that in almost every case the KR is that of a D-hydroxyl generating or B-type KR due to the presence of a diagnostic aspartate within the catalytic KR subdomain (Figure 5-6) (Keatinge-Clay, 2007; Piasecki et al., 2014). As DH domains eliminate B-type stereochemical products to *trans* olefins it seems that PS domains naturally act on *trans*-α/β unsaturated substrates. Ostensibly, SorPS9 seems an exception to this trend as the KR preceding it lacks the aforementioned diagnostic aspartate residue indicating it generates a β–D-hydroxy substrate that would give rise to a *cis* double bond after dehydration by the following DH. This anomaly is further supported by SorKR7's presence within the C-terminal end of split dehydrating bimodule previously shown to invariantly contain a KR_A in this position (Wagner et al., 2017). Between $SorKR_A7$ and

SorPS9 are two consecutive DH-like domains, the first of which almost certainly performs the elimination reaction generating the *cis*-olefin (Fig 5-1C). The following DH-like domain contains several aberrant features including a cysteine in place of the critical aspartate that normally completes the DH catalytic dyad and the lack of the conserved proline hypothesized to orient the catalytic histidine and thus more closely resembles a diene isomerizing domain (DH^0) rather than a traditional DH. A sequence alignment corroborates this assignment showing that the putative $SorDH^0$8 shares the most homology with a domain from the bonkregic acid pathway involved in the interconversion of an α/β-*cis*-γ/δ-*trans* with an α/β-*trans*-γ/δ-*cis* diene (Figure 5-7) (Moebius Chem Biol, Wagner Structure). This domain is bioinformatically distinct to the better characterized and seemingly more common enoyl-isomerase represented by PksEI16 that performs a α/β to β/γ double bond migration mechanism (Gay et al., 2014). It thus appears that in Nature, PS domains exclusively receive α/β-*trans*-unsaturated substrates and that at least in the case of SorPS9, the upstream isomerization machinery was adopted as a requisite to pyran ring installation.

To test the in vitro activity and selectivity of SorPS9 against α/β–unsaturated SNAC molecules we incubated the enzyme overnight with the anticipated PS substrate **1** predicted to form a 6-membered ring. After 16 hr. incubation at 25°C, LC-MS analysis of the reaction products revealed a near complete disappearance of **1** and the appearance of a new peak not observed in negative control reactions. Consistent with pyran formation, the new peak displayed an m/z ratio of 268 and a maximal absorbance shift to

234 nm from 260 nm as would be expected from the loss of α/β–conjugation with the thioester carbonyl that results from cyclization (Figure 5-8)**.** These results agree with the proposed role of SorPS9 in mediating pyran ring formation and further demonstrate the excised domain to be a robust and efficient *in vitro* biocatalyst. When SorPS9 was incubated with the cyclized product of **1**, no activity in the reverse direction was detected implying that equilibrium heavily favors the pyran ring formation (Figure 5-9).

To probe the substrate tolerance of SorPS9, we next challenged the enzyme with the racemic secondary alcohol substrate **2,** predicted to form a methyl substituted ring. The reaction was incubated as before with subsequent LC-MS analysis detecting two new peaks of roughly equivalent size. The appearance of these two equivalent peaks suggests that SorPS9 processes R and S secondary alcohols with a minimal preference for hydroxyl stereochemistry.

Owing to the size and unsaturation of the natural substrate, SorPS9 likely has a tolerance for substitutions at the E-carbon and also to points of unsaturation along the substrate. This combination of the previously demonstrated stereoselectivity with regard to ether bond formation and substrate promiscuity towards α/β–unsaturated SNAC compounds could make SorPS9 an attractive tool in the chemoenzymatic synthesis of natural products containing pyran or furan rings.

To better understand the roles of the two prominent conserved polar active site residues H33 and N186 in pyran ring formation the following point mutations were made: H33A, N186A, N186D and N186H. All mutants produced soluble protein and were incubated as before with **1** and **2** and analyzed via LC-MS (Figure 5-10 and 5-11).

For both substrates, no discernable activity was detected for the H33A mutant consistent with its assignment as the critical base necessary for proton abstraction. The N186H mutant retained significant but reduced activity supporting its possible role as hydrogen bond donor to the substrate nucleophilic hydroxyl. The N186A and N186D mutants each displayed minimal and no activity, respectively, further corroborating this hypothesis. Together, these mutants demonstrate that although N186 is not essential for catalysis it plays an important role in the reaction and seems to contribute greatly to the overall rate of cyclization. The complete abolition of activity observed for the N186D mutant further implies that N186 interacts with the substrate as a hydrogen bond donor since the introduction of negative charge in its place would be expected to create a thermodynamically unfavorable electronic interaction that would impede substrate binding. A hydrogen bond with the substrate oxygen makes sense in the context of this mechanism as it would stabilize the substrate transition state during proton abstraction but when removed as in the case of the N186A mutant would not necessarily extinguish catalytic activity.

The lack of potential catalytic residues within the SorPS9 active site allows for few possible players in the route to pyran ring formation. The SorPS9 structure and mutagenesis results in combination with previous functional and structural work on related enzymes strongly suggest that SorPS9 catalysis is primarily dependent upon H33 to mediate cyclization. From the information available, it is highly probable that H33 is the sole residue directly participating in catalysis, acting to activate the ζ-hydroxyl and subsequently donate the hydroxyl-abstracted proton to the polyketide α-position to

complete the Michael addition reaction (Figure 5-8). When substrate and product mimics are modeled into the SorPS9 active site cavity such that the ζ -hydroxyl is positioned in close proximity to H33 for proton abstraction the hydroxyl oxygen is within hydrogen binding proximity to N186 (~2.5 Å) (Fig 5-12B). Though not strictly required for catalysis, N186 also plays a strong supporting role to H33, likely by acting as a hydrogen bond donor to the polyketide substrate ζ-hydroxyl.

The high-resolution crystal structure of SorPS9 helps shed new light on the mysterious pyran synthase domains endogenous to many *trans*-AT polyketide synthases and highlights the high degree of homology PS domains share with the DH domain. Accompanying functional data demonstrate SorPS9 to be a robust *in-vitro* catalyst for the production of pyran rings with structure-guided mutagenesis implicating a single histidine as the only residue critical for catalysis. Structural comparison of PS to DH domains shows that slight alterations to DH active site architecture and overall secondary structure can lead to striking new types of chemistry. Indeed, modest alterations to the DH domain seem to have repeatedly repurposed it for new types of catalysis with structural and bioinformatic analysis suggesting it has served as a template for the evolution of novel PKS domains involved in polyketide cyclization, olefin shifting, *cis/trans* isomerization and vinylogous dehydration with at least one example of a tri-functional domain existing (Gay et al., 2014, Fiers et al., 2016; Wagner et al., 2017;, Berkhan et al., 2014).

The aberrant enzyme architecture, atypical domain ordering and novel chemistry within *trans*-AT synthases have hindered discernment of a coherent colinear relationship

between protein sequence and polyketide structure similar to what has been achieved for *cis*-AT systems. As the integration of structural, functional and bioinformatic research methodology continues to demystify the numerous idiosyncratic features of *trans*-AT synthases the intricate relationship between synthase and polyketide will more clearly emerge. To this end, the investigation of SorPS9 advances our understanding of the structures and chemistries hidden within the *trans*-AT realm and demonstrates how slight modifications to a malleable active site can expand the range chemistry performed by an enzyme family. Overall this work furthers the elucidation of *trans*-acyltransferase polyketide biosynthetic gene clusters that still possess a large untapped reservoir of biosynthetic potential.

METHODS

Cloning, Expression and Purification

The gene-region corresponding to SorPS9 was amplified from *Sorangium cellulosum* Soce 12 genomic DNA, Gibson Assembled (New England Biolabs) into the pET28b expression vector and expressed heterologously in *Escherichia coli* BL21(DE3). N-terminal-hexahistidine-tagged SorPS9 was purified from cell lysate using immobilized metal affinity followed by gel filtration. Active site mutants of SorPS9 were generated using Gibson Assembly and standard site-directed mutagenesis techniques and primers (Table S2). Resulting protein was purified in a similar manner to wild-type.

Crystallization and Structure Determination

Crystals of SorPS9 grew over 2–7 days by sitting drop vapor diffusion at 25 °C in a condition composed of 1.7 M ammonium sulfate and 0.1 M bis-tris (pH 5.5). The structure was initially solved by molecular replacement and the resulting model was further refined to 1.55 Å resolution using Arwarp, Refmac and Phenix Refine (Adams + Richardson, Langer et al., Winn et al., Murshudov et al.). The coordinates for SorPS9 have been deposited in the Protein Data Bank, accession code 6B2V.

Substrate Synthesis

1A and **1B** are commercially available (Figure 5-13).

Synthesis of **1A** and **1B** could be performed from the following protocol:

δ-valerolactone (1.0 g, 10.0 mmol) was dissolved in 30 ml dichloromethane, and then DIBAL-H 12 ml (1 M solution in hexane) was added at -78ºC. The mixture was kept string for 2 hours until it was done. 3 ml methanol and then 10 ml saturated potassium sodium tartrate solution was carefully added to quench the reaction. The mixture was extracted by EtOAc 50 ml twice, and dried over Na_2SO_4. Precipitation $Al(OH)_3$ could be removed by vacuum filtration. The solvent was removed by reduced pressure, and the crude **1A** (470 mg, 40.5 %) was directly used for the next step without further purification as a sticky colorless oil.

6-methyltetrahydro-2*H*-pyran-2-one (1.16 g, 10.0 mmol) was dissolved in 30 ml dichloromethane, and then DIBAL-H 12 ml (1 M solution in hexane) was added at -78ºC. The mixture was kept string for 2 hours until it was done. 3 ml methanol and then 10 ml

saturated potassium sodium tartrate solution was carefully added to quench the reaction. The mixture was extracted by EtOAc 50 ml twice, and dried over Na_2SO_4. Precipitation $Al(OH)_3$ could be removed by HCl wash. The solvent was removed by reduced pressure, and the crude mixture cis/trans-**1B** (248 mg, 21.4 %) was purified by column chromatography as a sticky colorless oil.

Column condition: 3x7 cm.

Eluant: 200 ml (9:1 EtOAc-hexane), then 100 ml pure EtOAc.

Chen, S.H. Hong, B.C. Su, C.F. Sarshar, S. (2005) Tetrahedron Letters, 46(51), 8899-8903.

Synthesis **2A**, **2B**, **2C**, and ***rac*-2D** (Figure 5-13) was adapted from the following protocol:

J. Kennedy, N. J. Mccorkindale, R. A. Raphael (1061) A new synthesis of queen substance. J. Chem. Soc. 1961, 3813-3815.

1A (470 mg, 4.60 mmol), malonic acid (530 mg, 5.09 mmol), 562 µl pyridine and 30 µl propylamine were added together, and kept stirring overnight. The reaction mixture was then subjected to oil bath 105ºC for 4 hours. The residue was poured into 2 ml 2M H_2SO_4 and extracted with diethyl ether 20ml and then 10 ml. The organic phase was dried out over Na_2SO_4. The solvent was removed by reduced pressure. Purification of the residue via column chromatography gave **2A** (171 mg, 25.7 %) as a sticky colorless oil. **2A** was described in the following protocol.

J. Kennedy, N. J. Mccorkindale, R. A. Raphael (1061) A new synthesis of queen substance. J. Chem. Soc. 1961, 3813-3815.

Column condition: 1x7 cm.

Eluant: 100 ml (1:1 EtOAc-hexane), then 100 ml pure EtOAc.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.06 (dt, *J* = 15.6, 6.9 Hz, 1H), 5.83 (dt, *J* = 15.6, 1.6 Hz, 1H), 3.66 (t, *J* = 6.1 Hz, 2H), 2.26 (qd, *J* = 7.1, 1.5 Hz, 2H), 1.72 – 1.46 (m, 4H) (Figure 5-14).

HRESIMS m/z 148.9 (M-H_2O+Na) (calcd for $C_7H_{10}O_2Na$ 149.06).

1B (248 mg, 2.14 mmol), malonic acid (245 mg, 2.13 mmol), 260 μl pyridine and 38 μl piperidine were added together, and kept stirring overnight. The reaction mixture was then subjected to oil bath 105ºC for 4 hours. The residue was poured into 2 ml 2M H_2SO_4 and extracted with diethyl ether 20ml and then 10 ml. The organic phase was dried out over Na_2SO_4. The solvent was removed by reduced pressure. Purification of the residue via column chromatography gave **2B** (144 mg, 27.2 %) as a sticky colorless oil.

NMR of **2B** was confirmed by previous reference: *J. Org. Chem.*, 1989, 54 (4), 911–914.

Column condition: 1x7 cm.

Eluant: 100 ml 1:1 EtOAc-hexane), then 100 ml pure EtOAc.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.07 (dt, *J* = 15.6, 6.9 Hz, 1H), 5.84 (dq, *J* = 15.6, 0.9 Hz, 1H), 3.82 (q, *J* = 6.1 Hz, 1H), 2.26 (td, *J* = 6.9, 1.4 Hz, 2H), 1.67 – 1.54 (m, 1H), 1.57 – 1.42 (m, 3H), 1.30 – 1.17 (m, 3H) (Figure 5-15).

HRESIMS m/z 163.1 (M-H_2O+Na) (calcd for $C_8H_{12}O_2Na$ 163.07).

1A (138 mg, 1.35 mmol), malonic acid (155 mg, 1.49 mmol), 166 μl pyridine and 10 μl piperidine were added together, and kept stirring overnight. The reaction mixture was then subjected to oil bath 105ºC for 4 hours. The residue was poured into 4 ml 1M HCl and extracted with diethyl ether 20 ml then 10 ml. The organic phase was dried out over Na_2SO_4. The solvent was removed by reduced pressure. Purification of the residue via column chromatography gave mainly **2C** (33 mg, 17.0 %) as a sticky slightly yellow oil. NMR of **2C** was confirmed by previous reference: Synthetic Communications, 1988, 18(15), 1757-62.

Column condition: 1x8 cm.

Eluant: 90 ml (1:2 EtOAc-hexane).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 4.00 (dt, *J* = 11.5, 2.2 Hz, 1H), 3.78 – 3.68 (m, 1H), 3.48 (td, *J* = 11.5, 2.9 Hz, 1H), 2.54 (dd, *J* = 15.5, 8.0 Hz, 1H), 2.46 (dd, *J* = 15.5, 4.8 Hz, 1H), 1.84 (ddd, *J* = 11.3, 5.7, 2.3 Hz, 1H), 1.71 – 1.45 (m, 4H), 1.41 – 1.28 (m, 1H) (Figure 5-16).

HRESIMS m/z 145.1 (M+H) (calcd for $C_7H_{13}O_3$ 145.09).

1B (573 mg, 4.94 mmol), malonic acid (557 mg, 5.36 mmol), 580 μl pyridine and 32 μl piperidine were added together, and kept stirring overnight. The reaction mixture was then subjected to oil bath 105ºC for 4 hours. The residue was poured into 2 ml 8M 1M HCl and extracted with diethyl ether 25ml twice. The organic phase was dried out over Na_2SO_4. The solvent was removed by reduced pressure. Purification of the residue via column chromatography gave mainly ***rac*-2D** (215 mg, 27.5 %) as a sticky slightly yellow oil.

NMR of ***rac*-2D** was confirmed by previous reference: *T.L.*, 2016, 57 (1), 53-55.

Column condition: 1x8 cm.

Eluant: 90 ml (1:2 EtOAc-hexane).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.77 (dtt, *J* = 10.8, 5.4, 2.4 Hz, 1H), 3.55 (ddd, *J* = 11.1, 6.3, 1.9 Hz, 1H), 2.61 – 2.44 (m, 2H), 1.68 – 1.47 (m, 4H), 1.30 – 1.22 (m, 2H), 1.20 (dd, *J* = 6.2, 3.8 Hz, 3H) (Figure 5-17).

HRESIMS m/z 159.1 (M+H) (calcd for $C_8H_{15}O_3$ 159.09).

Synthesis **3A**, **3B**, **3C** and ***rac*-3D (**Figure 5-13**)** was adapted from the following protocol and the results were compatible as previously reported:

Pöplau, P. Frank, S. Morinaka, B.I. Piel, J. (2013) Angew. Chem. Int. Ed. 2013, 52, 13215-13218.

EDC-HCl (384 mg, 2.00 mmol) was added to **2A** (145 mg, 1.00 mmol) in 2.5 ml dichloromethane solution at 0ºC under protection of argon. The reaction mixture was kept stirring for 20 minutes at 0ºC. After that N-acetylcysteamine (143 mg, 1.20 mmol) and 4-DMAP (20 mg, 0.163 mmol) was added to the mixture. The reaction was then moved to room temperature stirring for 4 hours. After stirring the reaction is quenched by 5 ml water, and extracted by EtOAc 10 ml three times. The organic phase was washed with brine and dried over Na_2SO_4. The solvent was removed under reduced pressure. Purification of the residue via HPLC (r_t = 14.8 min, estimated yield 4.7 %) gave **3A** as a yellow oil.

HPLC condition: Beckman Ultrasphere™ dp 5 µl mmx25 cm, flow rate 2.5 ml/min, 0.0 - 15.0 min 98.0% water/ 2.0% ACN; 15.0 - 20.0 min 70.0 %/ 30.0 % CAN; 20.0 -21.0 min 100% ACN; 22.0 -25.0 min 98.0 % water/ 2.0 % ACN.

^{1}H NMR (500 MHz, Acetonitrile-d_3) δ 6.94 (dt, J = 15.5, 6.9 Hz, 1H), 6.56 (s, 1H), 6.21 (dt, J = 15.5, 1.5 Hz, 1H), 3.52 (q, J = 5.8 Hz, 2H), 3.32 (q, J = 6.5 Hz, 2H), 3.07 – 2.94 (m, 2H), 2.33 – 2.21 (m, 2H), 1.85 (s, 3H), 1.53 (tt, J = 10.6, 5.6 Hz, 4H) (Figure 5-18).

HRESIMS m/z 268.1 (calcd for $C_{11}H_{19}NO_3SNa$ 268.10).

EDC-HCl (384 mg, 2.00 mmol) was added to **2B** (158 mg, 1.00 mmol) in 2.5 ml dichloromethane solution at 0°C under protection of argon. The reaction mixture was kept stirring for 20 minutes at 0°C. After that N-acetylcysteamine (143 mg, 1.20 mmol) and 4-DMAP (20 mg, 0.163 mmol) was added to the mixture. The reaction was then moved to room temperature stirring for 4 hours. After stirring the reaction is quenched by 5 ml water, and extracted by EtOAc 10 ml three times. The organic phase was washed with brine and dried over Na_2SO_4. The solvent was removed under reduced pressure. Purification of the residue via HPLC (r_t = 16.9 min, estimated yield 44.4 %) gave **3B** as a yellow oil.

HPLC condition: Beckman Ultrasphere™ dp 5 µl mmx25 cm, flow rate 2.5 ml/min, 0.0 - 15.0 min 98.0% water/ 2.0% ACN; 15.0 - 20.0 min 70.0 %/ 30.0 % CAN; 20.0 -21.0 min 100% ACN; 22.0 -25.0 min 98.0 % water/ 2.0 % ACN.

^{1}H NMR (400 MHz, Chloroform-d) δ 6.92 (dt, J = 14.2, 6.8 Hz, 1H), 6.13 (d, J = 15.5 Hz, 1H), 6.02 (s, 1H), 3.83 – 3.78 (m, 1H), 3.47 (q, J = 5.8 Hz, 2H), 3.08 (t, J = 6.3 Hz,

2H), 2.26 – 2.20 (m, 2H), 1.98 (s, 3H), 1.68 – 1.56 (m, 2H), 1.55 – 1.38 (m, 4H), 1.19 (d, *J* = 6.2 Hz, 3H) (Figure 5-19).

HRESIMS m/z 282.1 (calcd for $C_{12}H_{21}NO_3SNa$ 282.11).

EDC-HCl (44 mg, 0.229 mmol) was added to **2C** (33 mg, 0.229 mmol) in 2 ml dichloromethane solution at 0°C. After that N-acetylcysteamine (40 mg, 0.336 mmol) and 4-DMAP (4 mg, 0.036 mmol) was added to the mixture. The reaction was then moved to room temperature stirring overnight. After stirring the reaction is quenched by 4 ml saturated $(NH_4)_2SO_4$ solution. The aqueous phase was extracted by dichloromethane 5 ml twice. The organic phase was combined, washed with saturated sodium bicarbonate and dried over Na_2SO_4. The solvent was removed under reduced pressure. Further purification of the fraction via HPLC (r_t = 18.0 min, estimated yield 31.0 %) gave mainly **3C** as a yellow oil.

HPLC condition: Beckman Ultrasphere™ dp 5 µl mmx25 cm, flow rate 2.5 ml/min, 0.0 - 15.0 min 98.0% water/ 2.0% ACN; 15.0 - 20.0 min 70.0 %/ 30.0 % CAN; 20.0 -21.0 min 100% ACN; 22.0 -25.0 min 98.0 % water/ 2.0 % ACN.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.99 – 3.91 (m, 1H), 3.79 (dddd, *J* = 10.7, 8.4, 4.5, 2.1 Hz, 1H), 3.50 – 3.38 (m, 3H), 3.05 (td, *J* = 6.2, 2.5 Hz, 2H), 2.79 (dd, *J* = 14.8, 8.4 Hz, 1H), 2.61 (dd, *J* = 14.7, 4.5 Hz, 1H), 1.97 (s, 3H), 1.87 – 1.80 (m, 1H), 1.55 (dtdd, *J* = 19.8, 16.8, 7.3, 5.3 Hz, 4H), 1.38 – 1.30 (m, 1H) (Figure 5-20).

HRESIMS m/z 268.1 (calcd for $C_{11}H_{19}NO_3SNa$ 268.10).

EDC-HCl (261 mg, 1.36 mmol) was added to ***rac*-2D** (215 mg, 1.36 mmol) in 10 ml dichloromethane solution at 0°C. After that N-acetylcysteamine (232 mg, 1.95 mmol) and

4-DMAP (24 mg, 0.214 mmol) was added to the mixture. The reaction was then moved to room temperature stirring overnight. After stirring the reaction is quenched by 8 ml saturated $(NH_4)_2SO_4$ solution. The organic phase was washed with brine and dried over Na_2SO_4. The solvent was removed under reduced pressure. The reaction mixture was first separated by a $CuSO_4$ contained silica-gel column. Further purification of the fraction via HPLC (r_t = 18.0 min, estimated yield 51.0 %) gave mainly ***rac*-3D** as a yellow oil.

Column condition: 1x5 cm.

Eluant: 60 ml (1:2 EtOAc-hexane), then 60 ml (1:1 EtOAc-hexane).

HPLC condition: Beckman Ultrasphere™ dp 5 µl mmx25 cm, flow rate 2.5 ml/min, 0.0 - 15.0 min 98.0% water/ 2.0% ACN; 15.0 - 20.0 min 70.0 %/ 30.0 % CAN; 20.0 -21.0 min 100% ACN; 22.0 -25.0 min 98.0 % water/ 2.0 % ACN.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.82 (s, 1H), 3.84 – 3.75 (m, 1H), 3.49 – 3.40 (m, 3H), 3.09 – 2.98 (m, 2H), 2.82 (dd, *J* = 14.8, 7.7 Hz, 1H), 2.67 – 2.53 (m, 1H), 1.96 (d, *J* = 5.0 Hz, 3H), 1.85 – 1.78 (m, 1H), 1.62 – 1.45 (m, 3H), 1.22 – 1.17 (m, 2H), 1.14 (d, *J* = 6.2 Hz, 3H) (Figure 5-21).

HRESIMS m/z 282.1 (calcd for $C_{12}H_{21}NO_3SNa$ 282.11).

Activity assays

All enzyme reactions were incubated with 15 mM substrate overnight in 150 mM NaCl and 20 mM HEPES pH 7.5 at 5 mg/mL enzyme concentration. After incubation, samples were 3X extracted in ethyl-acetate, evaporated and brought up in 50 μL methanol prior to being submitted for LC-MS analysis (Agilent 6120 Quadrupole ESI in the positive mode equipped with an Agilent 1260 HPLC system).

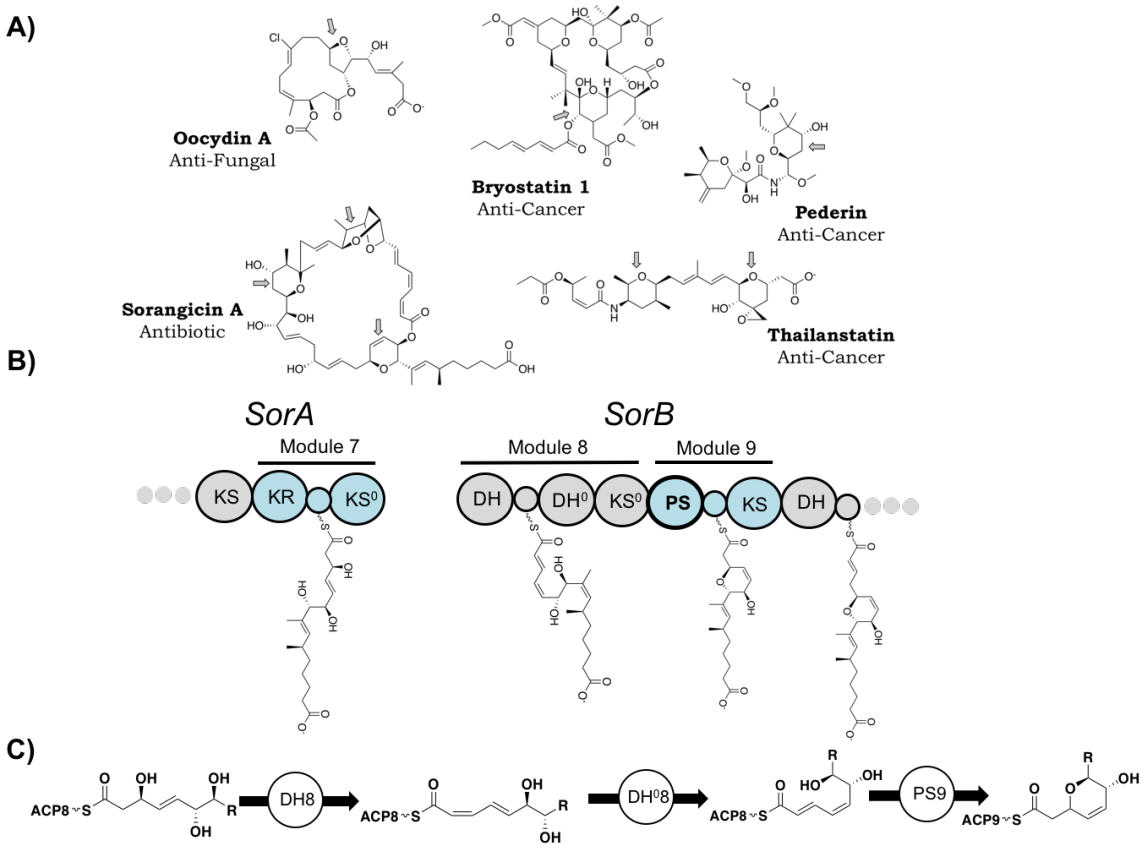


Figure 5-1: Pyran synthase domains in *trans*-acyltransferase polyketide synthases.

(A) Selected trans-AT polyketides containing pyran synthase derived pyran rings. Gray arrows denote pyran rings generated by embedded pyran synthase domains.

(B) Domain organization of modules 6-8 of the sorangicin synthase. KS = ketosynthase, KR = ketoreductase, KS^0 = condensation incompetent ketosynthase, DH = dehydratase, DH^0 = olefin isomerase, PS = pyran synthase.

(C) Domain activity immediately preceding SorPS9.

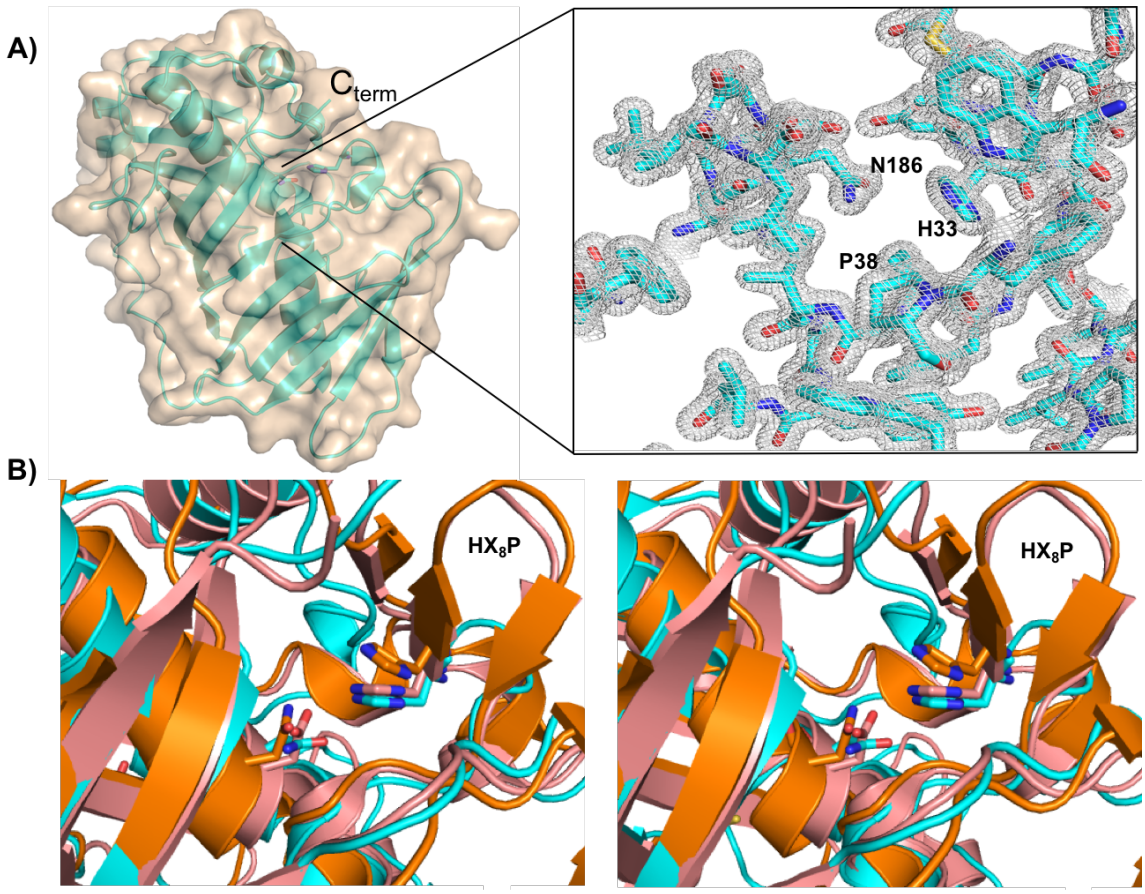


Figure 5-2: Structure of SorPS9.

(A) Overall structure of SorPS9 (left). $2F_0$-F_C omit map of SorPS9 active site conserved active site residues.

(B) Stereodiagram of SorPS9 active site (cyan) overlayed with *trans*-AT dehydratase domain RzxB (Salmon - PDB: 5IL6) and *trans*-AT enoyl-isomerase domains PksEI16 (orange – PDB:4U3V).

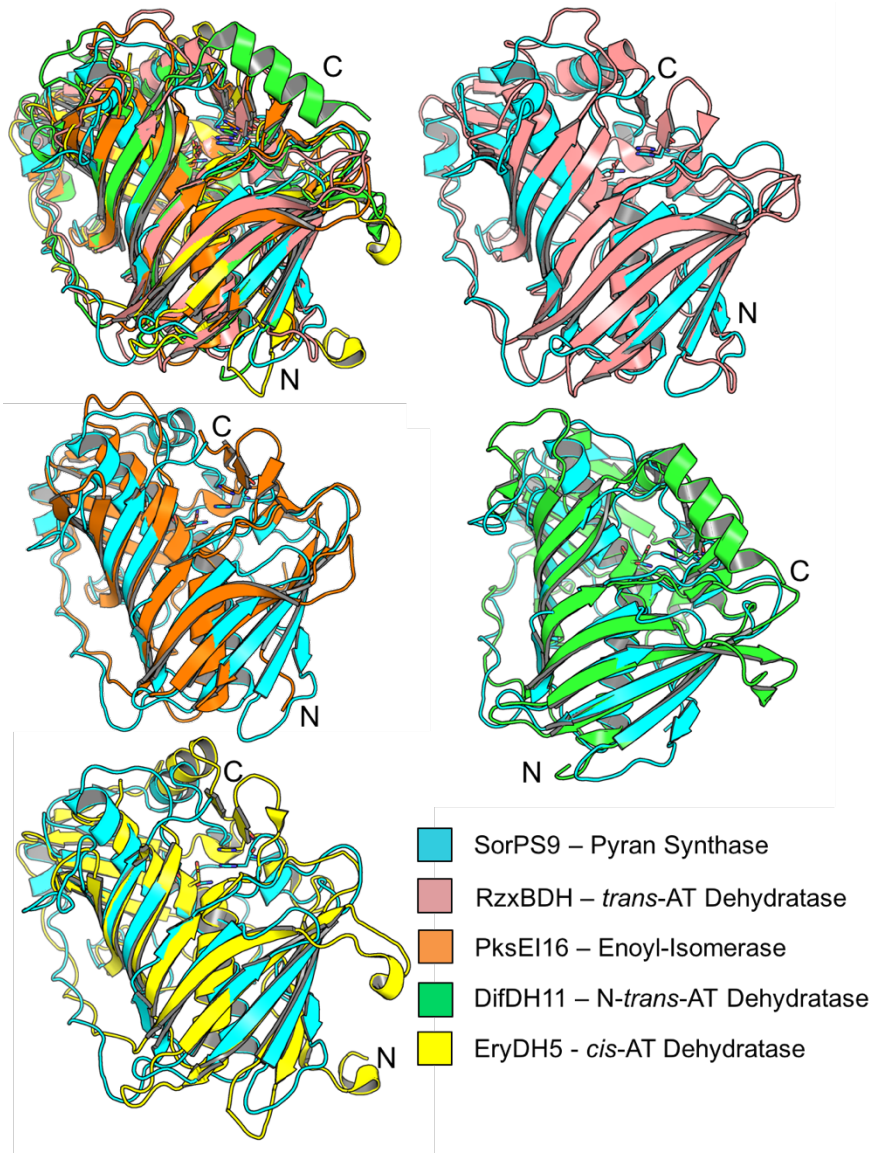


Figure 5-3: Structural overlay of SorPS9 with homologous domains.

Figure 5-3 continued;

Comparison of SorPS9 (teal: PBD 6B2V) with *trans*-AT dehydratase RzxBDH (salmon: PDB 5IL6), Enoyl-isomerase PksEI16 (orange: PDB 3U3V) N-terminal dehydratase DifDH11 (green: PDB 4KKU) and *cis*-AT dehydratase EryDH5 (yellow: PDB 3IL6). In addition to mutation and rearrangement of catalytic dyad residues, structural variation between these domains is concentrated around the N-terminus as well as the β– sheets behind the active site histidine residue.

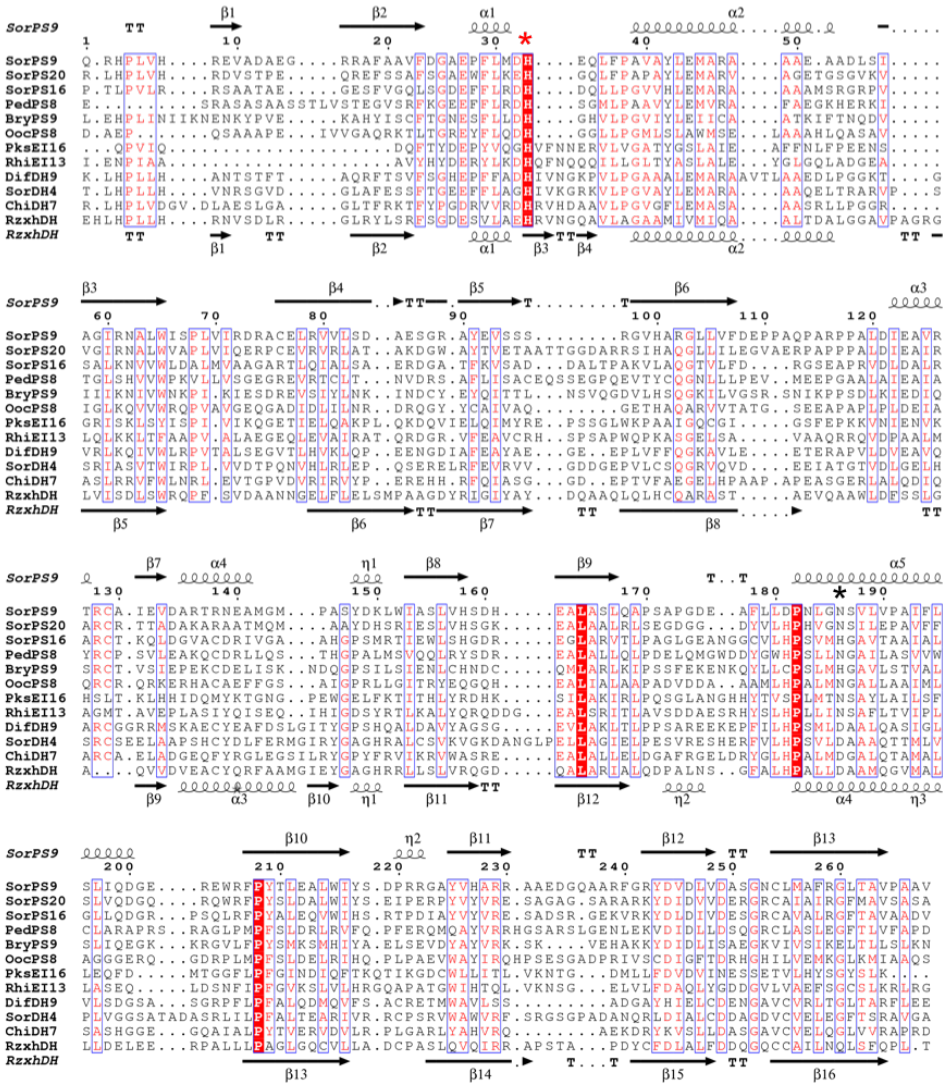


Figure 5-4: Pyran synthase sequence alignment.

Sequence and structural comparison of pyran synthase (PS), dehydratase (DH) and enoyl-isomerase (EI) domains. Top depicts secondary structure of SorPS9, bottom depicts secondary structure of *trans*-AT DH RzxhDH (PDB: 5IL6). Red asterisk indicates

Figure 5-4 continued;

universally conserved catalytic histidine. Black asterisk indicates additional active residue completing catalytic dyad: N/H in PS, D in DH, N in EI.

Figure 5-5: Multiple sequence of trans-AT pyran synthase domains.

Figure 5-5 continued;

Common domain alterations compared to the dehydratase progenitor include a truncation to the HX_8P motif that follows the catalytic histidine (red asterisk) and substitution of DH active site aspartate, generally for an asparagine.

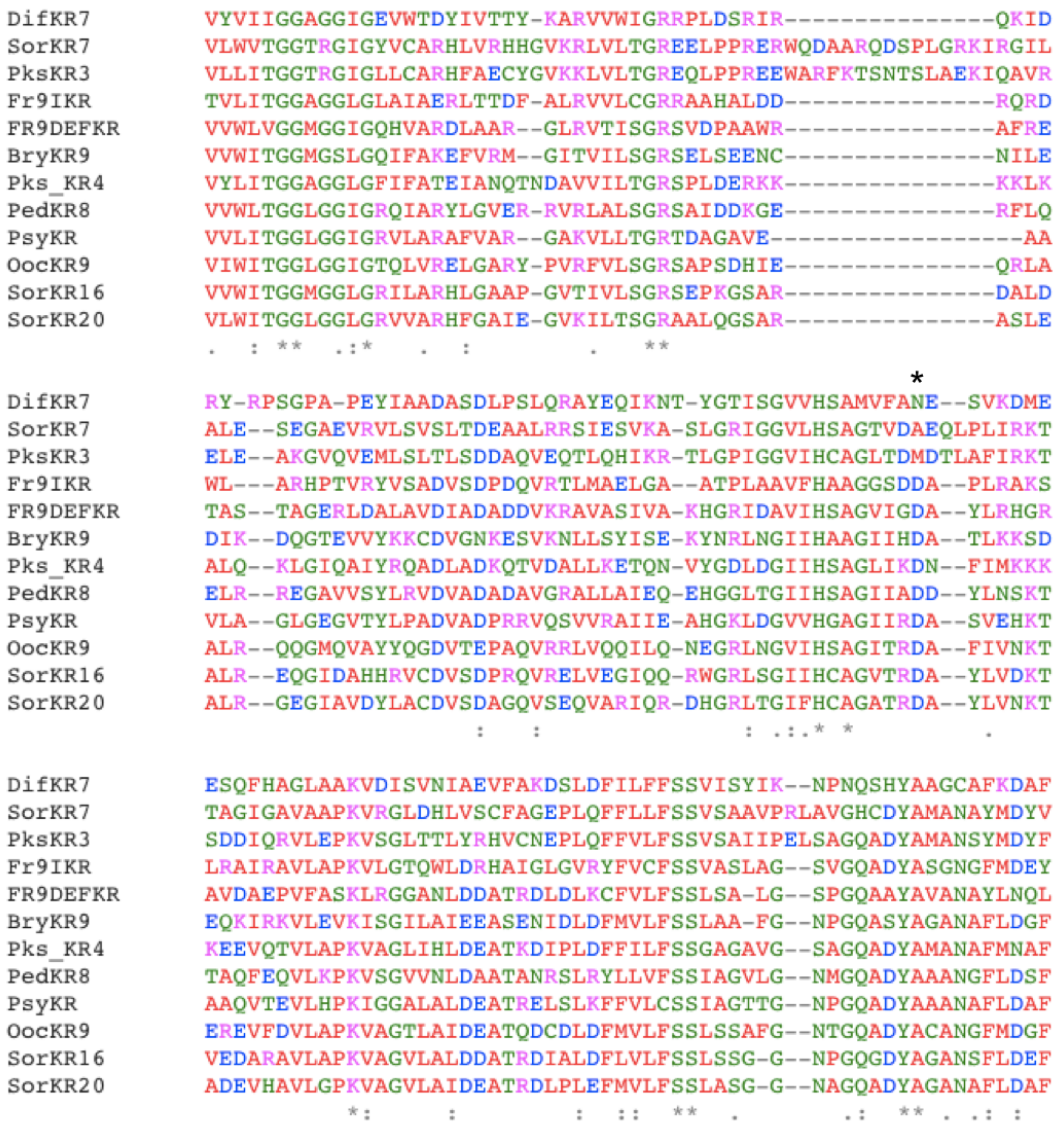


Figure 5-6: Clustal-omega alignment of PS preceeding KR structural subdomains. Asterisk denotes diagnostic aspartate indicative of B-type ketoreduction. Dif_KR4(A-type), Pks_KR3(A-type) and Pks4 (B-type) included as reference, all other sequences precede a PS domain.

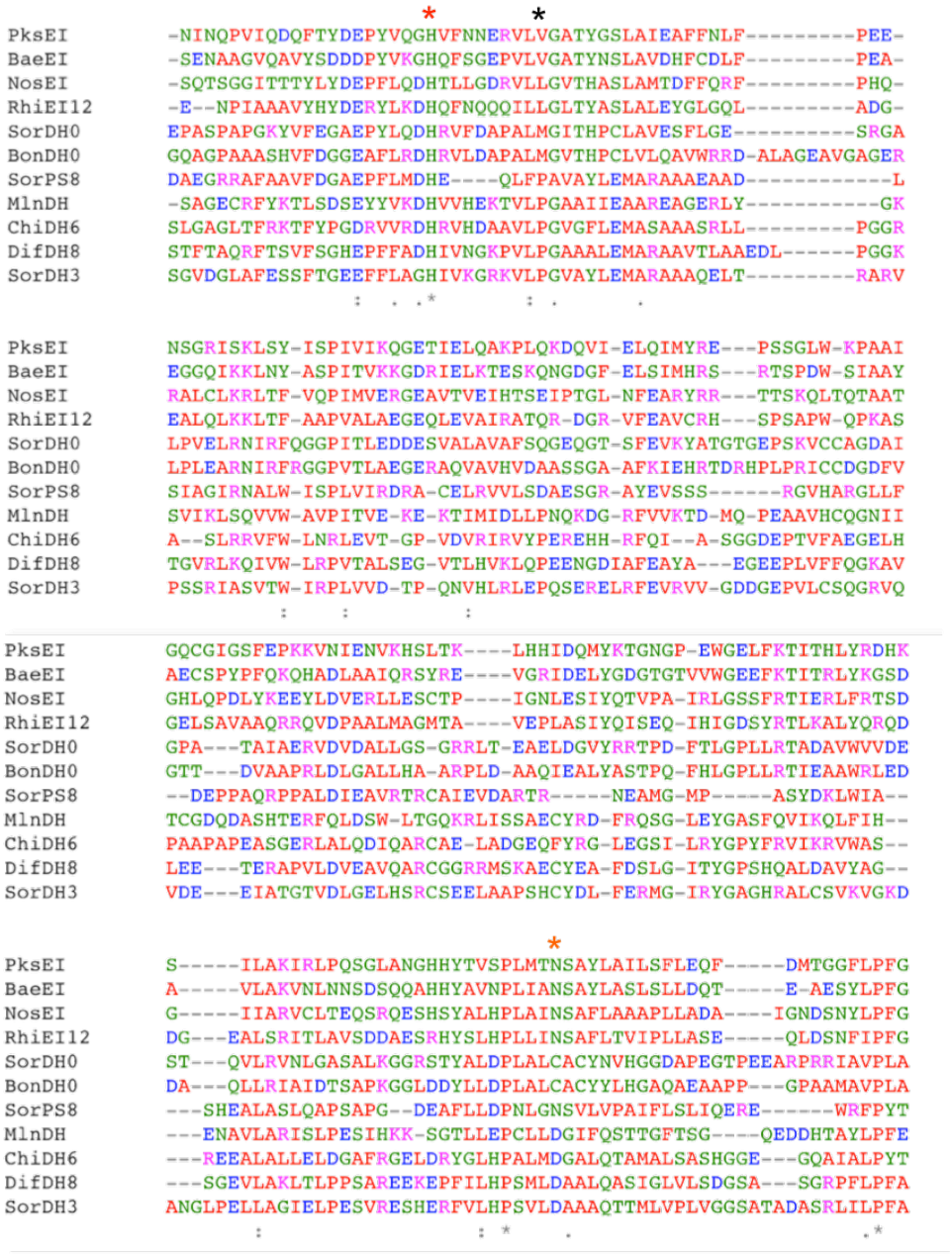


Figure 5-7: Clustal-omega alignment of PS, DH and EI domains.

Clustal-omega pyran synthase (PS), dehydratase(DH), di-ene-isomerase(DH^0) and enoyl-isomerase (EI) domains. Red asterisk denotes active site histidine. Black asterisk denotes

Figure 5-7 continued;

conserved proline of DH HX_8P motif. Orange asterisk denotes catalytic dyad residue (aspartate in DH, histidine/asparagine in PS, asparagine in EI, cysteine in DH^0).

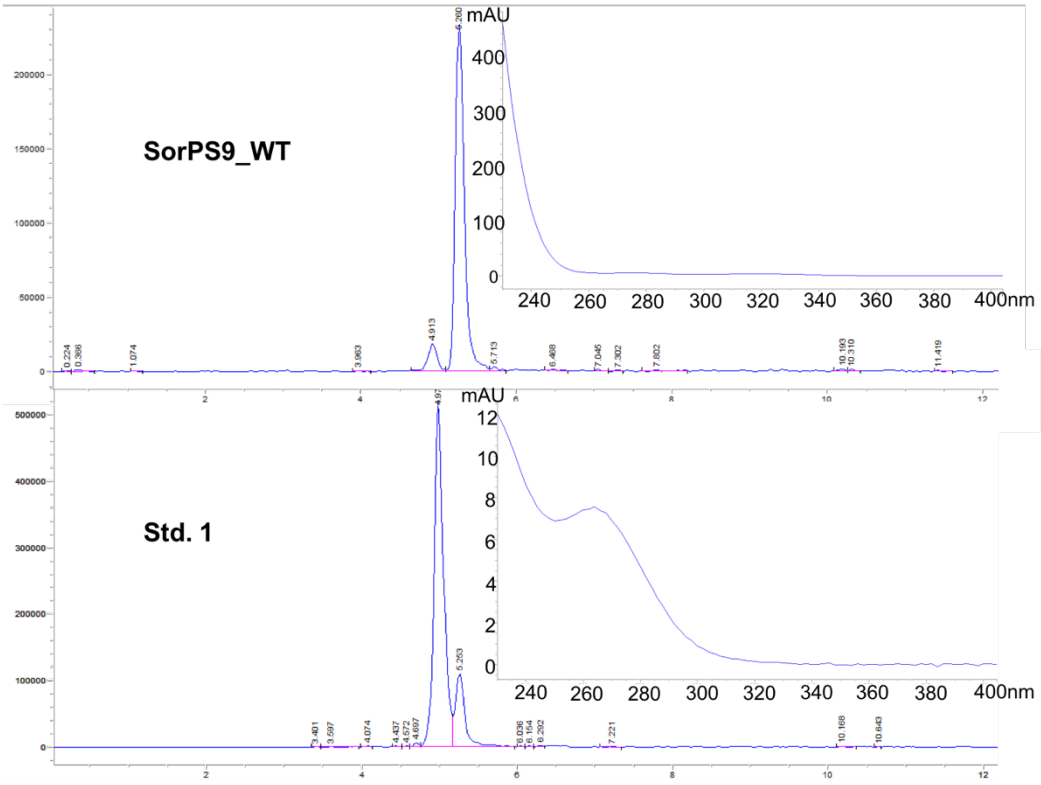


Figure 5-8: Absorbance characteristics of SorPS9 reaction products.

After incubation with SorPS9, the absorbance profile of the product shifted compared with that of substrate **1** from 268 nm to 234 nm.

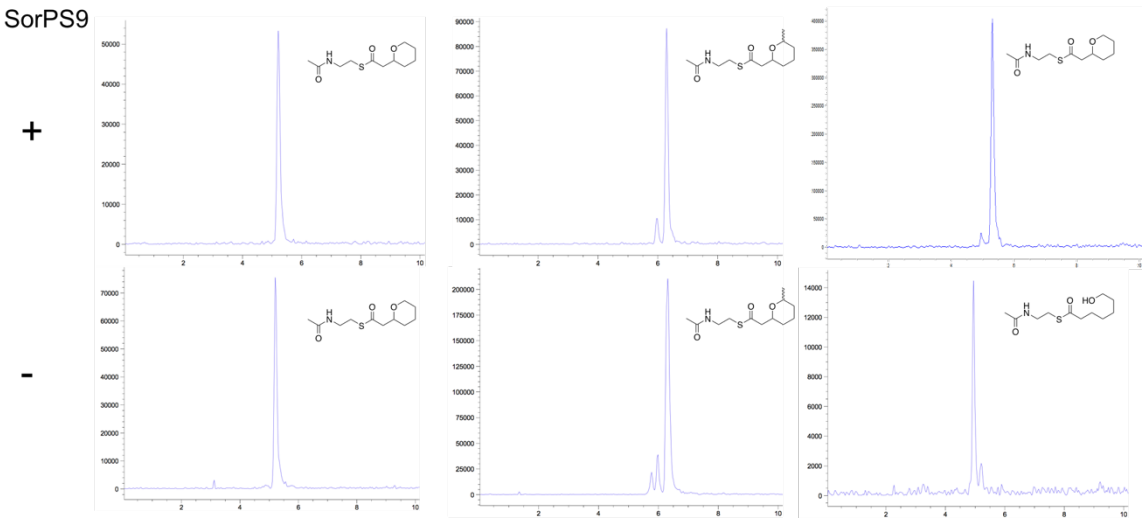


Figure 5-9: SorPS9 reverse reaction.

SorPS9 was incubated with unsubstituted pyran, substituted pyran, and unsubstituted substrate (left middle and right columns, respectively) with conversion only being observed for the reaction moving in the biosynthetic direction.

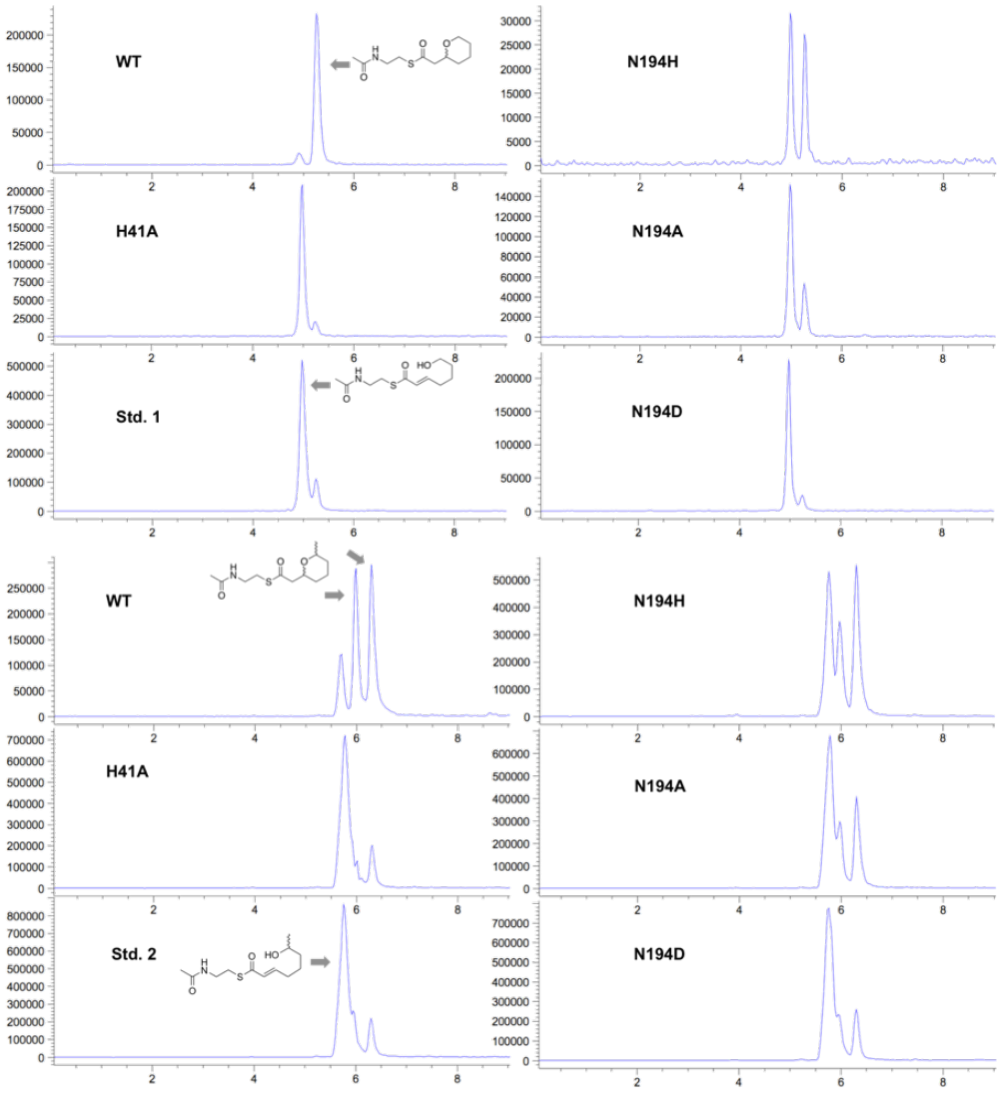


Figure 5-10: SorPS9 activity assays and mutagenesis.

LC-MS analysis of SorPS9 and active site mutants with (*E*)-7-hydroxyhept-2-enethioc-SNAC (left) and racemic (*E*)-7-hydroxyoct-2-enethioic-SNAC (right) with extracted m/z of 268 and 282, respectively.

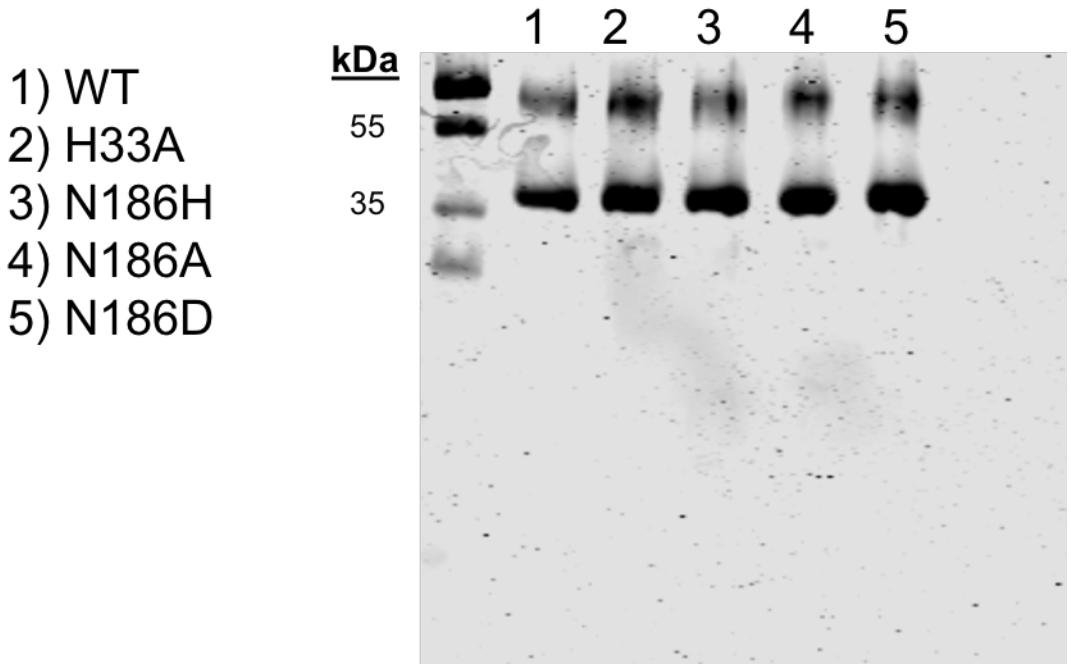


Figure 5-11: SDS-PAGE gel of SorPS9 and active site mutants.

Protein was purified via nickel-affinity chromatography followed by gel-filtration into low salt buffer. MW of SorPS9 = 31 KDa.

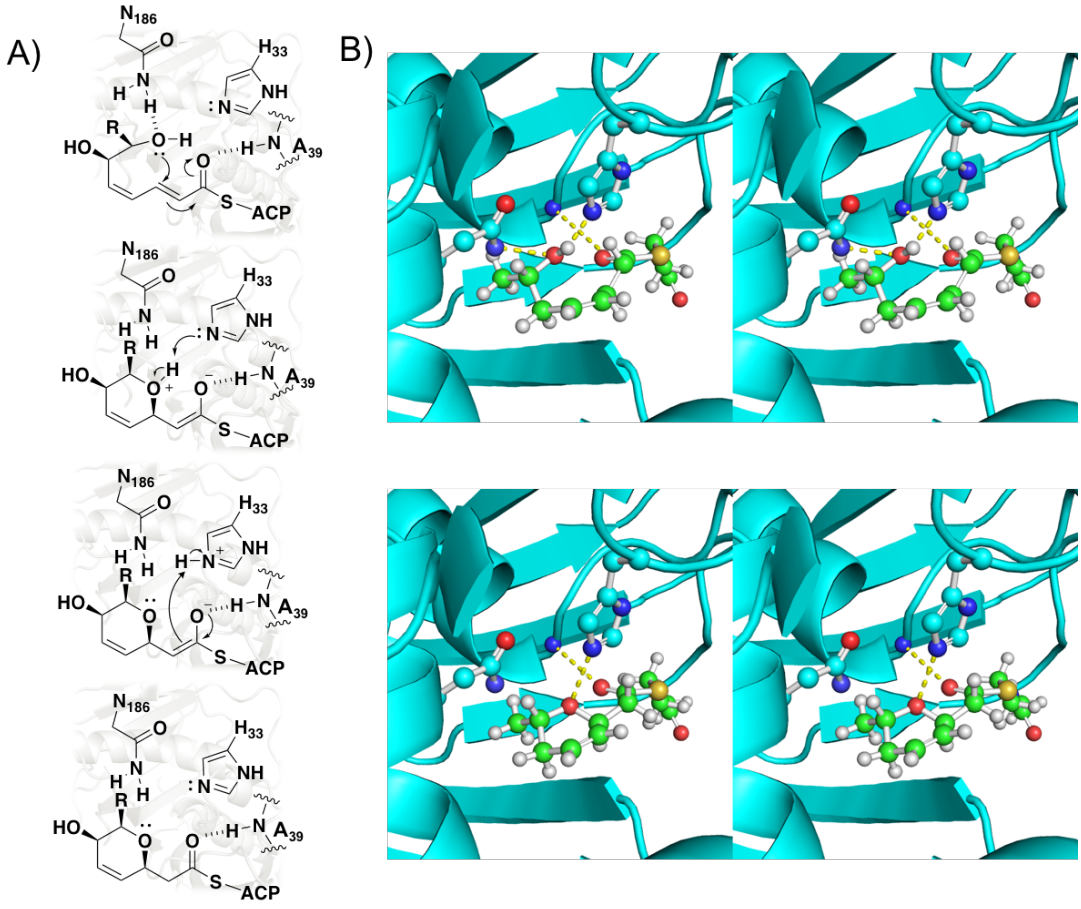


Figure 5-12: Proposed Mechanism of Pyran Ring Formation.

(A) Proposed mechanism of SorPS9 pyran ring formation. C-O bond formation is facilitated by H-bond of N186 with substrate hydroxyl prior to proton abstraction by H33. Backbone amide of A39 creates oxy-anion hole to stabilize thioester polarization. Subsequent collapse of oxy-anion intermediate regenerates SorPS9 active site.

(B) Modelling of substrate and product mimics into active highlights probable key ligand interactions with His33, Asn186 and backbone amide of A39.

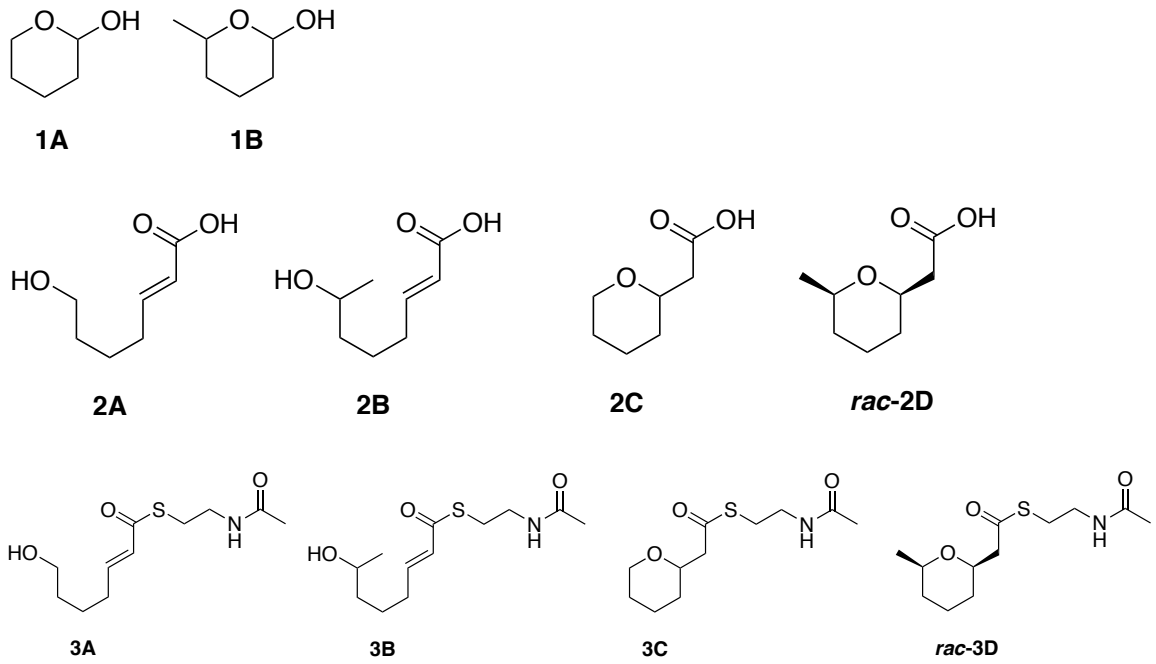


Figure 5-13: Pyran synthase substrate reagents.

Reagents (**1A** and **1B**) used in synthesis of pyran synthase substrates and standards (**2A-2D** and **3A-3D**). Refer to Substrate Synthesis section of Methods section for further information.

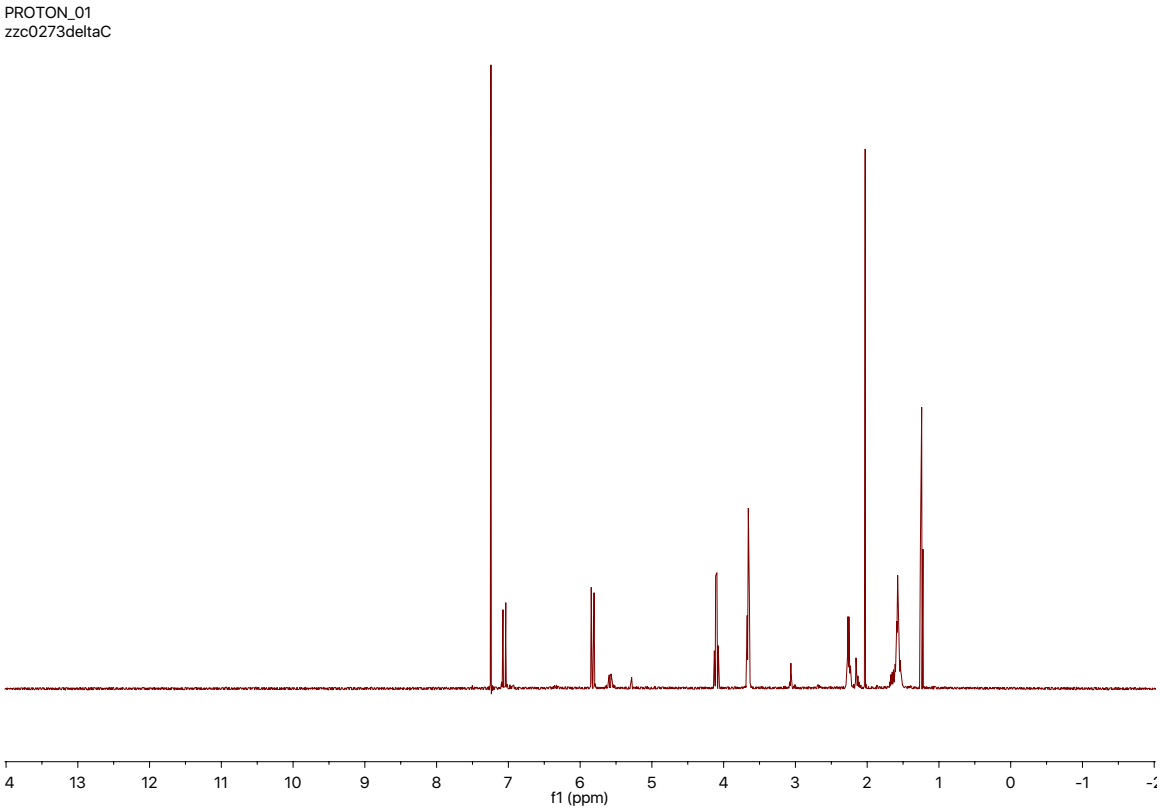


Figure 5-14: ^{1}H NMR of substrate **2A**.

(400 MHz, Chloroform-*d*) δ 7.06 (dt, *J* = 15.6, 6.9 Hz, 1H), 5.83 (dt, *J* = 15.6, 1.6 Hz, 1H), 3.66 (t, *J* = 6.1 Hz, 2H), 2.26 (qd, *J* = 7.1, 1.5 Hz, 2H), 1.72 – 1.46 (m, 4H).

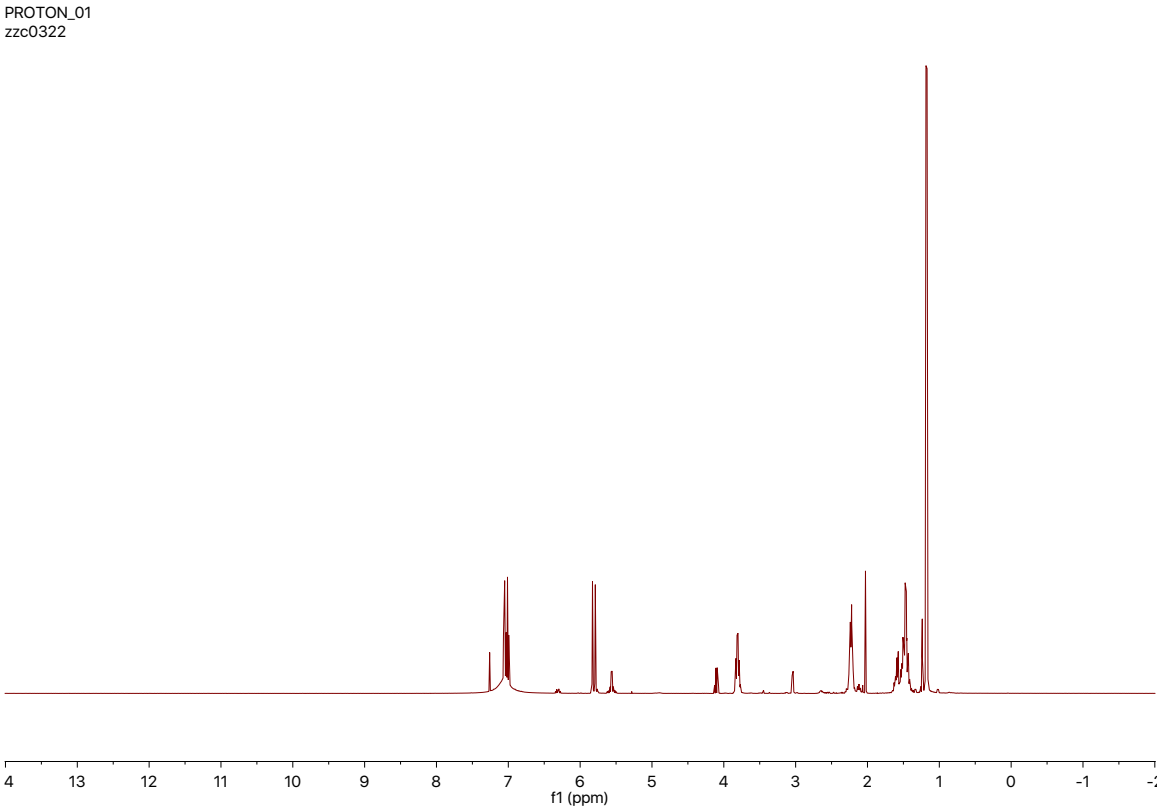


Figure 5-15: ^{1}H NMR of substrate **2B**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.07 (dt, *J* = 15.6, 6.9 Hz, 1H), 5.84 (dq, *J* = 15.6, 0.9 Hz, 1H), 3.82 (q, *J* = 6.1 Hz, 1H), 2.26 (td, *J* = 6.9, 1.4 Hz, 2H), 1.67 – 1.54 (m, 1H), 1.57 – 1.42 (m, 3H), 1.30 – 1.17 (m, 3H).

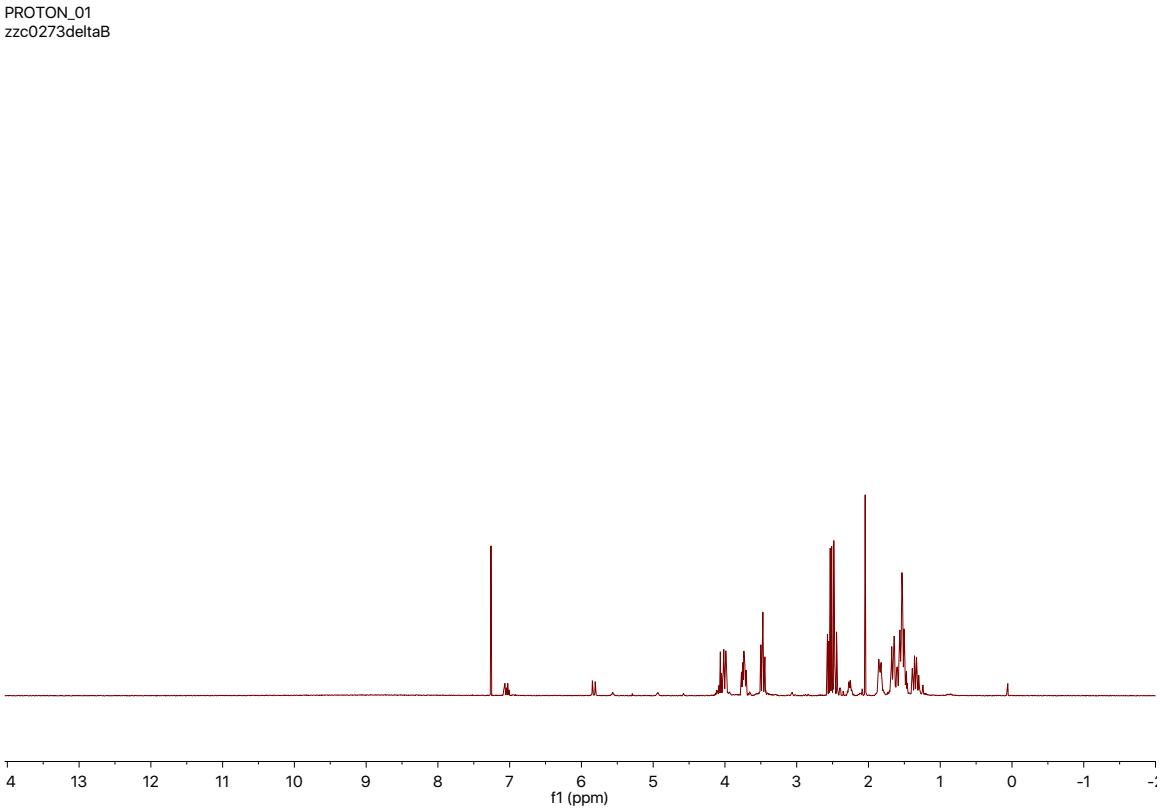


Figure 5-16: ^{1}H NMR of substrate **2C**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 4.00 (dt, *J* = 11.5, 2.2 Hz, 1H), 3.78 – 3.68 (m, 1H), 3.48 (td, *J* = 11.5, 2.9 Hz, 1H), 2.54 (dd, *J* = 15.5, 8.0 Hz, 1H), 2.46 (dd, *J* = 15.5, 4.8 Hz, 1H), 1.84 (ddd, *J* = 11.3, 5.7, 2.3 Hz, 1H), 1.71 – 1.45 (m, 4H), 1.41 – 1.28 (m, 1H).

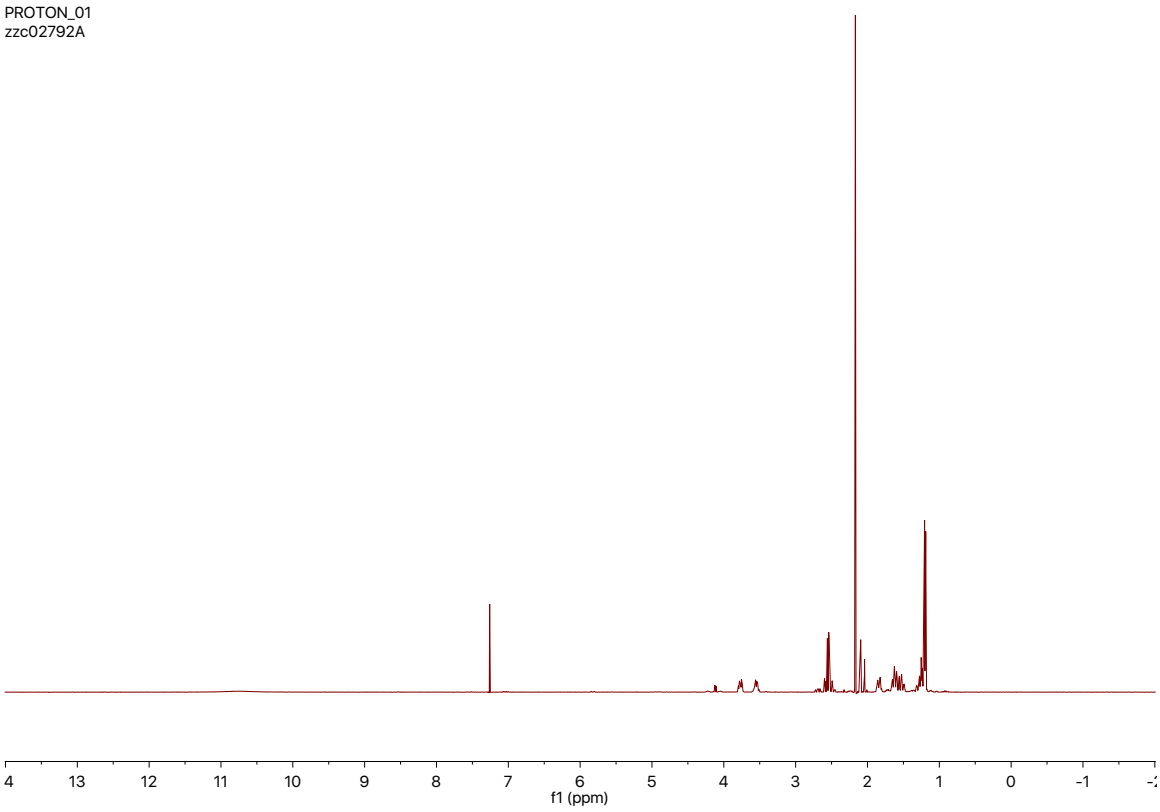


Figure 5-17: ^{1}H NMR of substrate ***rac*-2D**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.77 (dtt, *J* = 10.8, 5.4, 2.4 Hz, 1H), 3.55 (ddd, *J* = 11.1, 6.3, 1.9 Hz, 1H), 2.61 – 2.44 (m, 2H), 1.68 – 1.47 (m, 4H), 1.30 – 1.22 (m, 2H), 1.20 (dd, *J* = 6.2, 3.8 Hz, 3H).

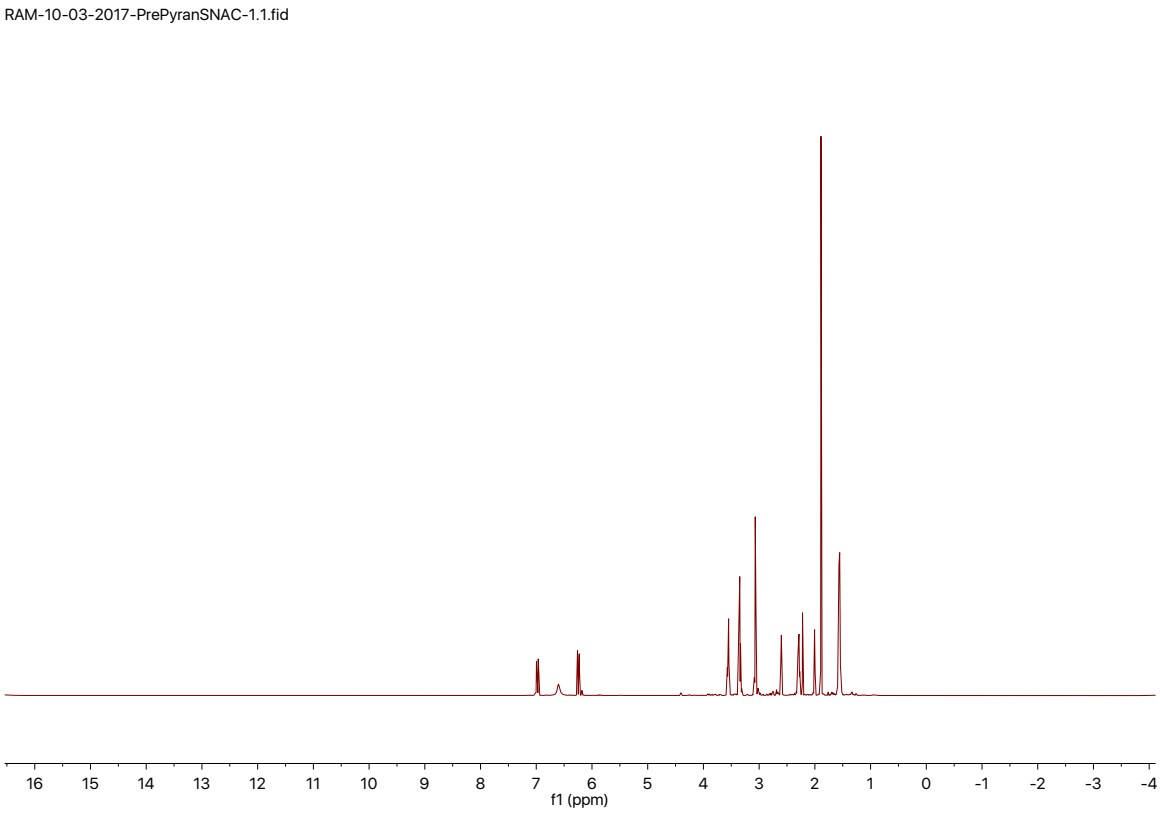


Figure 5-18: ^{1}H NMR of substrate **3A**.

^{1}H NMR (500 MHz, Acetonitrile-d_3) δ 6.94 (dt, J = 15.5, 6.9 Hz, 1H), 6.56 (s, 1H), 6.21 (dt, J = 15.5, 1.5 Hz, 1H), 3.52 (q, J = 5.8 Hz, 2H), 3.32 (q, J = 6.5 Hz, 2H), 3.07 – 2.94 (m, 2H), 2.33 – 2.21 (m, 2H), 1.85 (s, 3H), 1.53 (tt, J = 10.6, 5.6 Hz, 4H).

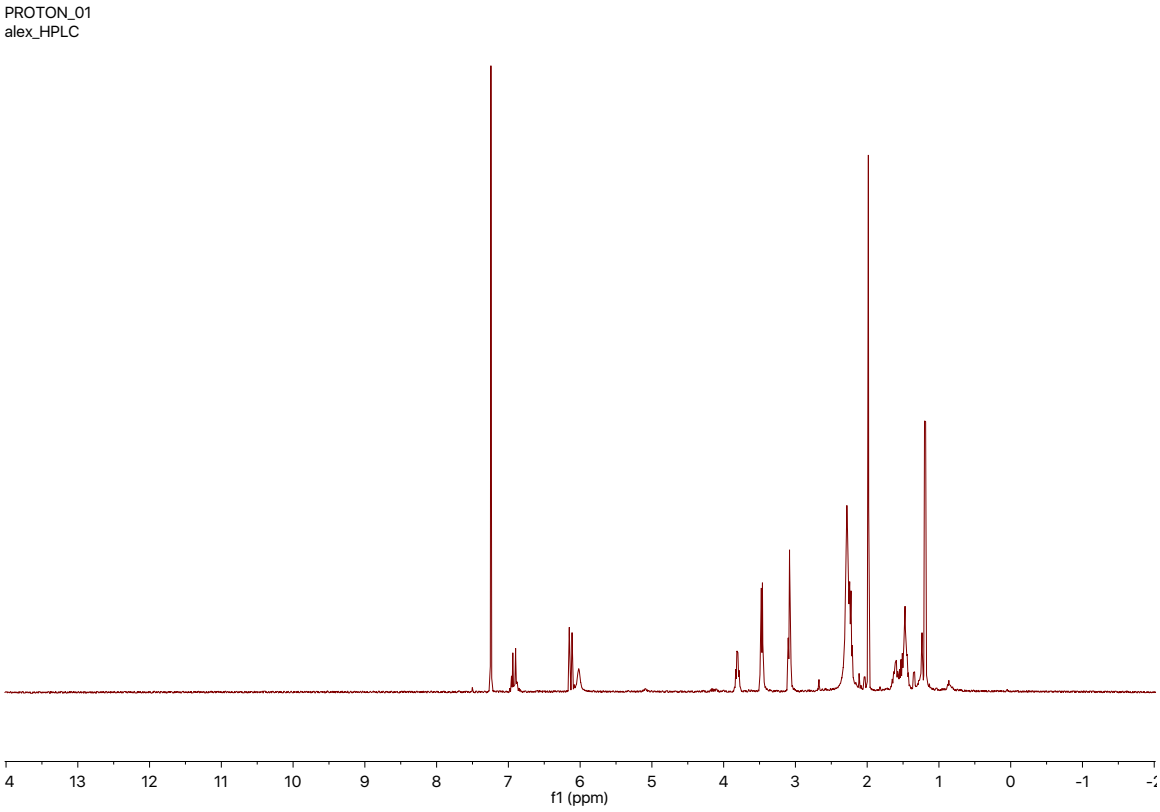


Figure 5-19: ^{1}H NMR of substrate **3B**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.92 (dt, *J* = 14.2, 6.8 Hz, 1H), 6.13 (d, *J* = 15.5 Hz, 1H), 6.02 (s, 1H), 3.83 – 3.78 (m, 1H), 3.47 (q, *J* = 5.8 Hz, 2H), 3.08 (t, *J* = 6.3 Hz, 2H), 2.26 – 2.20 (m, 2H), 1.98 (s, 3H), 1.68 – 1.56 (m, 2H), 1.55 – 1.38 (m, 4H), 1.19 (d, *J* = 6.2 Hz, 3H)

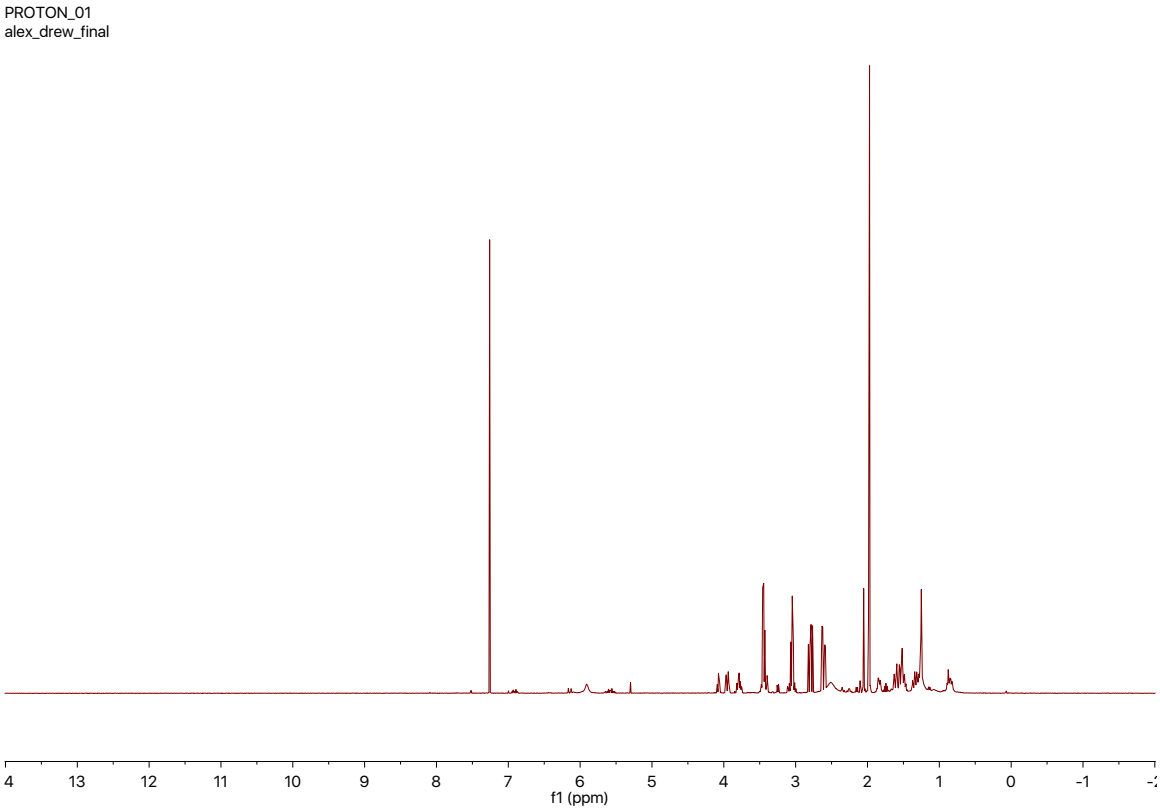


Figure 5-20: ^{1}H NMR of substrate **3C**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.99 – 3.91 (m, 1H), 3.79 (dddd, *J* = 10.7, 8.4, 4.5, 2.1 Hz, 1H), 3.50 – 3.38 (m, 3H), 3.05 (td, *J* = 6.2, 2.5 Hz, 2H), 2.79 (dd, *J* = 14.8, 8.4 Hz, 1H), 2.61 (dd, *J* = 14.7, 4.5 Hz, 1H), 1.97 (s, 3H), 1.87 – 1.80 (m, 1H), 1.55 (dtdd, *J* = 19.8, 16.8, 7.3, 5.3 Hz, 4H), 1.38 – 1.30 (m, 1H).

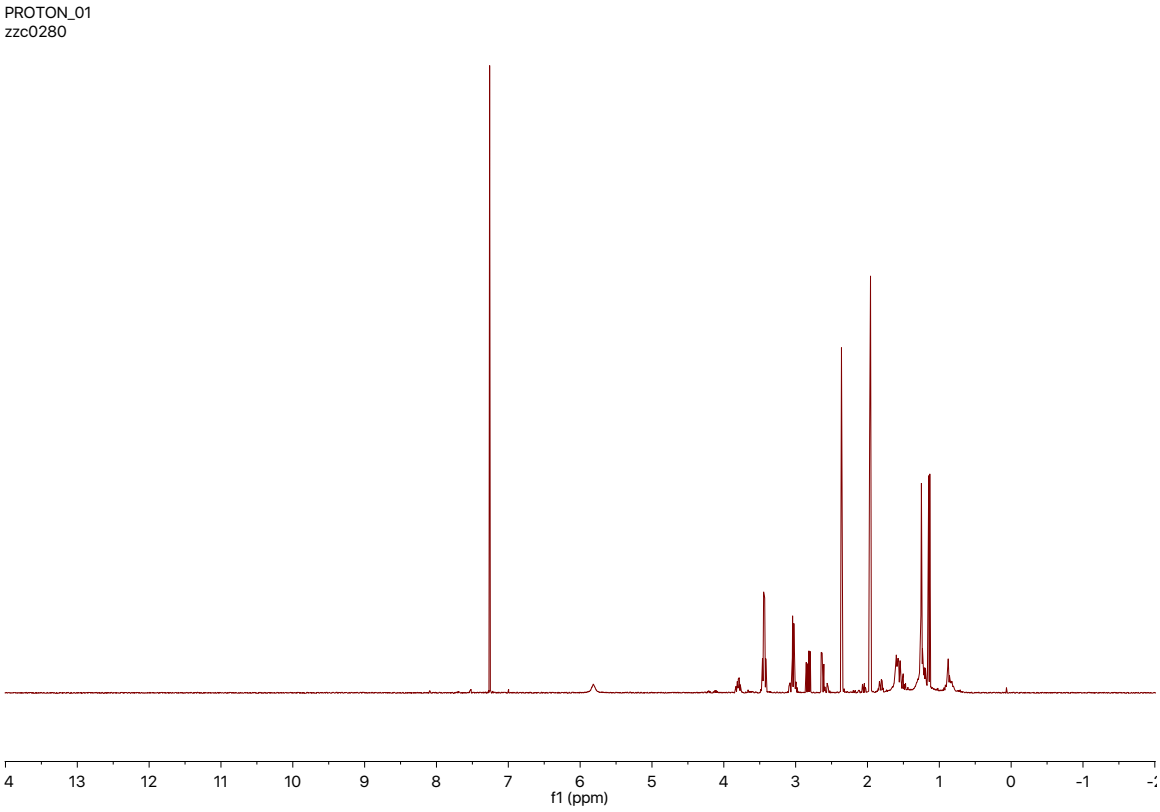


Figure 5-21: ^{1}H NMR of substrate **3D**.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.82 (s, 1H), 3.84 – 3.75 (m, 1H), 3.49 – 3.40 (m, 3H), 3.09 – 2.98 (m, 2H), 2.82 (dd, *J* = 14.8, 7.7 Hz, 1H), 2.67 – 2.53 (m, 1H), 1.96 (d, *J* = 5.0 Hz, 3H), 1.85 – 1.78 (m, 1H), 1.62 – 1.45 (m, 3H), 1.22 – 1.17 (m, 2H), 1.14 (d, *J* = 6.2 Hz, 3H).

Data collection	SorPS8
Wavelength (Å)	**1.0332**
Space group	**P 61 2 2**
Cell dimensions *a, b, c* (Å)	**75.37 75.37 195.87**
Resolution (Å)	**30.96 - 1.55 (1.61 - 1.55)**
R_{merge}	**0.064(0.422)**
$I/\sigma(I)$	**37.1(2.7)**
No. of reflections	**47028 (4385)**
Completeness (%)	**96.75 (91.97)**
Redundancy	**18.8(13.9)**
Wilson B value (Å^2)	**16.27**
Refinement	
Resolution (Å)	**1.55**
No. of reflections	**47024 (4385)**
R_{work}/R_{free}	**0.200/0.211**
No. of atoms	
Protein	**2057**
Water	**305**
Average *B* factors (Å^2)	
Protein	**21.2**
Water	**32.71**
RMS deviations	
Bond lengths (Å)	**0.028**
Bond angles (°)	**1.91**
Ramachandran Statistics (%)	
Preferred Regions	**96.98**
Allowed Regions	**2.64**
Outliers	**0.38**
Clashscore	**3.2**

Values in parentheses are for highest-resolution shell.

Table 5-1: SorPS9 data collection and refinement statistics.

DW469	SorPS9_F	gcctggtgccgcgcggcagccatATGGGAGAGGTGCGCCAGC
DW470	SorPS9_R	tctcagtggtggtggtggtggtgtcaCGCCGCTGGCTTCTCGG
DW504	SorPS9_H33A_F_GibsonSDM	CGGAGCCTTTCCTGATGGAT***GC***CGAGCAGCTGTTTCCGGCGG
DW505	SorPS9_H33A_R_GibsonSDM	CCGCCGGAAACAGCTGCTCG***GC***ATCCATCAGGAAAGGCTCCG
DW506	SorPS9_N186A_F_GibsonSDM	TCGACCCGAACCTGGGC***GC***TTCGGTTCTGGTGCCC
DW507	SorPS9_N186A_R_GibsonSDM	GGGCACCAGAACCGAA***GC***GCCCAGGTTCGGGTCGA
DW508	SorPS9_N186D_F_GibsonSDM	CCTCGACCCGAACCTGGGC***GA***TTCGGTTCTGGTGCCCGC
DW509	SorPS9_N186D_R_GibsonSDM	GCGGGCACCAGAACCGAA***TC***GCCCAGGTTCGGGTCGAGG
DW511	SorPS9_N186H_F_GibsonSDM	TCGACCCGAACCTGGGC***CA***TTCGGTTCTGGTGCCC
DW512	SorPS9_N186H_R_GibsonSDM	GGGCACCAGAACCGAA***TG***GCCCAGGTTCGGGTCGA

Table 5-2. Primers used in SorPS9 cloning and mutagenesis.

Lower case nucleotides indicate Gibson overlap regions. Bold italicized nucleotides indicate site directed mutagenesis regions.

Chapter 6: Conclusion

This work has explored the architecture and function of some of the less commonly occurring enzymes within type I polyketide synthases to help build a more complete picture of the detailed chemistry utilized by Nature for polyketide biosynthesis. This work is part of the continued basic research into these powerful enzymes which will ultimately provide the understanding necessary for their competent rational programming and engineering custom polyketide production.

Chapter 2 focused on the activity of embedded methyltransferase domains from *cis*-AT polyketide synthases. The work involved in vitro assays of six excised monofunctional MT domains from the gephyronic acid biosynthetic pathway using a panel of ACP bound and unbound thioester substrates. The methylation activity of these domains was monitored by HPLC and mass spectrometry and unequivocally showed that they operate exclusively on β-keto-acyl substrates and not on malonyl substrates as had been previously reported for two other *cis*-AT PKS methyltransferases. The work presents compelling evidence that methylation occurs only after KS mediated condensation for all canonical MTs within the gephyronic acid pathway, a finding that is likely to be general for the majority of other *cis*-AT MTs.

Chapter 3 built upon the work of **Chapter 2** by investigating methyltransferase domains from systems within the poorly understood *trans*-AT PKS bifurcation. MTs from 3 synthases: bacillaene, difficidin and macrolactin were shown to exclusively

methylate the α-position of β-keto-acyl substrates as was previously shown for MTs from *cis*-AT systems. Michaelis-Menten kinetic assays were subsequently performed with the aforementioned substrates revealing modest kinetic values for the six MTs assayed. Slow turnover and Km values in the mM range are not unusual when dealing with isolated PKS domains. Despite this, over 72 hr. time scales the MTs converted the majority of substrate to the anticipated methylated product.

Chapter 4 investigated the split bimodules common among the *trans*-AT synthases. These module splits commonly conform to a bimodular architecture wherein a module is split between a C-terminal KS domain and an N-terminal DH domain. A combination of structural biology, bioinformatics and enzymology was used to examine the aberrant features that are highly conserved among the split bimodular architecture. Structures of KR and DH domain from two split bimodules are presented that exemplify the novel structural motifs found with the split dehydrating bimodules providing insight into the finer details of the idiosyncrasies of the trans-AT synthases. In contrast to their more organized *cis*-AT counterpart, *trans*-AT synthases exhibit extensive structural and functional deviation including non-canonical chemistry, inactive domains, *trans*-acting components and as is the subject of this chapter modules split across adjacent polypeptides.s

In Chapter 5, a novel type of *trans*-AT PKS domain, the pyran synthase, responsible for the formation of 5 and 6 membered oxygen heterocycles during polyketide elongation is characterized. As the primary mechanism for pyran ring

formation within *trans-AT* polyketide synthases, pyran synthase domains are found within many polyketide biosynthetic gene clusters. The work presented in Chapter 5 reveals the PS domain to be closely related to the canonical DH domain and shows how a conserved DH active site histidine has been repurposed for new functionality. The high-resolution structure of the pyran synthase domains from the ninth module of the sorangicin along with structure guided mutagenesis of key residues suggests a mechanism wherein a conserved asparagine stabilizes the oxygen nucleophile of the substrate for activation by the histidine leading to ring closure. The structure demonstrates how subtle changes to an enzyme can transform it for new catalysis and furthers our understanding of atypical chemistry within *trans*-AT pathways.

We have unraveled many of the details concerning the chemistry of classical PKS pathways such as that seen in the erythromycin synthase but much mystery still surrounds the aberrant enzyme architecture and novel chemistry used in polyketide biosynthesis, particularly with regard to trans-AT pathways. This lack of understanding hinders elucidation of the collinear relationship between synthase and bioactive product that makes polyketide systems such attractive engineering targets. This work and future work like it will hone our understanding of the enigmatic features ok PKS biology and in so doing aid future efforts for engineering these marvelous enzymatic assembly lines for production of novel bioactive agents.

References

1) Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Nature. 348 176–78.

2) Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L. (1991) Modular organization of genes required for complex polyketide biosynthesis. Science. 252 675–79.

3) Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE. (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase. Annu Rev Biochem. 76 195-221.

4) Ansari, M. Z. Sharma, J. Gokhale, S. and Mohanty, D. (2008) In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinf. 9, 454.

5) Liscombe, D., Louie, G.V. Noel, J. P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases. Nat. Prod. Rep. 29, 1238-1250.

6) Ishida, K. Fritzsche, K. and Hertweck, C. (2007) A Gene Cluster Encoding Resistomycin Biosynthesis in Streptomyces resistomycificus. Exploring Polyketide Cyclization beyond Linear and Angucyclic Patterns J Am Chem Soc. 42, 129.

7) Schenk, Z. Xu, C. Pfeiffer, C. Steinbeck, and C. Hertweck. (2007) Geminal Bismethylation Prevents Polyketide Oxidation and Dimerization in the Benastatin Pathway. Angew Chem Int Ed. 46, 7035.

8) Williams, G. (2013) Engineering polyketide synthases and nonribosomal peptide synthetases. Curr Opin Struct Biol. 23, 603.

9) Keatinge-Clay, A.T. (2012) The structures of type I polyketide synthases. Nat. Prod. Rep. 29, 1050-1073.

10) Winter, J. M. Chiou, G. Bothwell, I.R. Xu, W. Garg, K.G. Luo, KM. Tang, Y. (2013) Expanding the Structural Diversity of Polyketides by Exploring the Cofactor Tolerance of an Inline Methyltransferase Domain. Org Lett. 15, 3775.

11) Miller, D. A. Luo, L. Hillson, N. Keating, T. A., and Walsh, C. T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. Chem. Biol. 9, 333-344.

12) Poust, S. Phelan, R.M. Deng, K. Katz, L. Petzold, C.J. Keasling, J.D. (2015) Divergent Mechanistic Routes for the Formation of gem-Dimethyl Groups in the Biosynthesis of Complex Polyketides. Angew

Chem Int Ed. 54, 2370.

13) Young, J. Stevens, D.C. Carmichael, R. Tan, J. Rachid, S. Boddy, C.N. Müller, R. Taylor, R.E. (2013) Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations. J Nat Prod. 76, 2269.

14) Hendricks, C.L. Ross, J.R. Pichersky, E. Noel, J.P.. Zhou, Z.S. (2004) An enzyme-coupled colorimetric assay for S-adenosylmethionine-dependent methyltransferases. Anal Biochem. 326, 100.

15) Cacho, R.A. Thuss, J. Xu, W. Sanichar, R. Gao, Z. Nguyen, A. Vederas, J.C. Tang, Y. (2015) Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. J Am Chem Soc. 137, 15688.

16) Gay, G. Wagner, D.T. Keatinge-Clay, A.T. Gay, D.C. (2014) Rapid modification of the pET-28 expression vector for ligation independent cloning using homologous recombination in Saccharomyces cerevisiae. Plasmid. 7, 66-71.

17) Piasecki, S.K. Taylor, C.A. Detelich, J.F. Liu, J. Zheng, J. Komsoukaniants, A. Siegel, D.R. (2011) Keatinge-Clay, A.T. Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. Chem. Biol. 18, 1331.

18) Piel, J. (2010) Biosynthesis of polyketides by trans-AT polyketide synthases. Nat.Prod. Rep. 27, 996-1047.

19) Helfrich, E. J., Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. Nat. Prod. Rep. 33, 231-316.

20) Cacho, R.A. Thuss, J. Xu, W. Sanichar, R. Gao, Z. Nguyen, A. Vederas, J.C. Tang, Y. (2015) Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. J. Am. Chem. Soc. 137, 15688-15691.

21) Jenner, M. Frank, S. Kampa, A. Kohlhaas, C. Pöplau, P. Briggs, G.S. Piel J. Oldham, N.J. (2013) Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. Angew. Chem. Int. Ed. 52, 1143-1147.

22) Jenner, M. Afonso, J.P., Bailey, H.R. Frank, S. Kampa, A. Piel, J. Oldham, N.J. (2015) Acyl-chain elongation drives ketosynthase substrate selectivity in trans-acyltransferase polyketide synthases. Angew. Chem. Int. Ed. 54, 1817-1821.

23) Moldenhauer, J. Götz, D.C. Albert, C.R. Bischof, S.K. Schneider, K. Süssmuth, R.D. Engeser, M. Gross, H. Bringmann, G. and Piel, J. (2010) The final steps of bacillaene biosynthesis in Bacillus amyloliquefaciens FZB42: direct evidence for beta,gamma dehydration by a trans-acyltransferase polyketide synthase. Angew. Chem. Int. Ed. Engl. 49, 1465-1467.

24) Chen, X.H. Vater, J. Piel, J. Franke, P. Scholz, R. Schneider, K. Koumoutsi, A. Hitzeroth, G. Grammel, N. Strittmatter, A.W. Gottschalk, G. Süssmuth, R.D. Borriss, R. (2006) Structural and

functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. J. Bacteriol. 188, 4024-4036.

25) Gurney, R. and Thomas, C. M. (2011) Mupirocin: biosynthesis, special features and applications of an antibiotic from a gram-negative bacterium. Appl. Microbiol.Biotechnol. 90, 11-21.

26) Bailey, C.B. Pasman, M.E. and Keatinge-Clay, A. T. (2015) Substrate structure-activity relationships guide rational engineering of modular polyketide synthase ketoreductases. Chem. Commun. (Camb). 52, 792-795.

27) Johnson, K. A. Simpson, Z. B. and Blom, T. (2009) Global Kinetic Explorer: A newcomputer program for dynamic simulation and fitting of kinetic data. Analytical Biochemistry 387, 20-29.

28) Johnson, K. A. Simpson, Z. B. and Blom, T. (2009) FitSpace Explorer: An algorithm to evaluate multi-dimensional parameter space in fitting kinetic data. Analytical Biochemistry 387, 30-4.

29) Johnson, K. A. (2009) Fitting enzyme kinetic data with KinTek global kinetic explorer. Computer Methods in Enzymology 467, 601-626.

30) Nguyen, T. Ishida, K. Jenke-Kodama, H. Dittman, E., Gurgui, G. Hochmuth, T. Taudien, S. Platzer, M. Hertweck, C. and Piel, J. (2008). Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. 26, 225-233.

31) Siskos, A.P. Baerga-Ortiz, A. Bali, S. Stein, V. Mamdani, H. Spiteller, D. Popovic, B. Spencer, J.B. Staunton, J. Weissman, K.J. Leadlay, P.F. (2005) Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. Chem. Biol. 12, 1145-1153.

32) Li, Y. Fiers, W.D. Bernard, S.M. Smith, J.L. Aldrich, C.C. Fecik, R.A. (2014) Polyketide intermediate mimics as probes for revealing cryptic stereochemistry of ketoreductase domains. ACS Chem. Biol. 9, 2914-22.

33) Wang, M. And Boddy, C. N. (2008) Examining the role of hydrogen bonding interactions in the substrate specificity for the loading step of polyketide synthase thioesterase domains. Biochemistry. 47, 11793-11803.

34) Horsman, M. E. Hari, T. P. and Boddy, C. N. Polyketide synthase and non- ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? (2016) Nat. Prod. Rep. 33, 183-202.

35) Li, Y. Dodge, G.J. Fiers, W.D. Fecik, R.A. Smith, J.L. Aldrich, C.C. (2015) Functional Characterization of a Dehydratase Domain from the Pikromycin Polyketide Synthase. J. Am. Chem. Soc. 137, 7003-7006.

36) Zheng, J. Gay, D. C. Demeler, B., White, M.A. and Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. (2012) Nat. Chem. Biol. 8, 615-621.

37) Akey, D.L. Razelum, J.R. Teransia, J. Sherman, D.H. Gerwick, W.H.

and Smith, J.L. (2010) Crystal structures of dehydrase domains from the curacin polyketide biosynthetic pathway. Structure 18, 94-105.

38) Amagai, K. Takaku, R. Kudo, F. and Eguchi, T. (2013). A unique amino transfer mechanism for constructing the β-amino fatty acid starter unit in the biosynthesis of the macrolactam antibiotic cremimycin. Chembiochem 14, 1998-2006.

39) Bailey, C.B., Pasman, M.E., and Keatinge-Clay, A.T. (2016). Substrate structure-activity relationships guide rational engineering of modular polyketide synthase ketoreductases. Chem. Commun. 52, 792-795.

40) Bonnett, S.A. Whicher, J.R. Papireddy, K. Florova, G. Smith, J.L. and Reynolds, K.A. (2013). Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. Chem. Biol. 20, 772-783.

41) Bretschneider, T. Heim, J.B. Heine, D. Winkler, R. Busch, B. Kusebauch, B. Stehle, T. Zocher, G. and Hertweck, C. (2013). Vinylogous chain branching catalysed by a dedicated polyketide synthase module. Nature 502, 124-128.

42) Butcher, R.A. Schroeder, F.C. Fischbach, M.A. Straight, P.D. Kolter, R. Walsh, C.T. and Clardy, J. (2007). The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. Proc. Natl. Acad. Sci. USA 104, 1506-1509.

43) Caffrey, P. (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. ChemBioChem 4, 654–657.

44) Cheng, Y.Q. Tang, G.L. and Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. USA 100, 3149-3154.

45) Dorival, J. Annaval, T. Risser, F. Collin, S. Roblin, P. Jacob, C. Gruez, A. Chagot, B., and Weissman, K.J. Characterization of intersubunit communication in the virginiamycin trans-acyltransferase polyketide synthase. 2016. J. Am. Chem. Soc. 138, 4155-4167.

46) Emsley, P. Lohkamp, B. Scott, W.G. and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486-501.

47) Gay, D.C. You, Y.O. Keatinge-Clay, A.T. and Cane, D.E. (2013). Structure and stereospecificty of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. Biochemistry 52, 8916-8928.

48) Gay, D.C. Gay, G. Axelrod, A.J. Jenner, M. Kohlhaas, C. Kampa, A. Oldham, N.J. Piel, J. and Keatinge-Clay, A.T. (2014). A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. Structure 22, 444-451.

49) Gay, D.C. Spear, P.J. and Keatinge-Clay, A.T. (2014). A double-hotdog fold with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl isomerase. ACS Chem. Biol. 17, 2374-2381.

50) Haines, A.S. Dong, X. Song, Z. Farmer, R. Williams, C. Hothersall, J.

Poskon, E. Wattana-Amorn, P. Stephens, E.R. Yamada, E. Gurney, R. Takebayashi, Y. Masschelein, J. Cox, R.J. Lavigne, R. Willis, C.L. Simpson, T.J. Crosby, J. Winn, P.J. Thomas, C.M. Crump, M.P. (2013) A conserved motif flags acyl carrier proteins for beta-branching in polyketide synthesis. Nat. Chem. Biol. 9, 685-692.

51) Hertweck, C. (2010). The biosynthetic logic of polyketide diversity. Angew. Chem. Int. Ed. 48, 4688-4716.

52) Hur, G.H. Vickery, C.R. and Burkart M.D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Nat. Prod. Rep. 29, 1074-1098.

53) Irschik, H. Kopp, M. Weissman, K.J. Buntin, K. Piel, J., and Müller, R. (2010). Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. Chembiochem 11, 1840-1849.

54) Keatinge-Clay, A.T., and Stroud, R.M. (2006). The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. Structure 14, 737-748.

55) Keatinge-Clay, A.T. (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. 14, 898–908.

56) Keatinge-Clay, A.T. (2008). Crystal structure of the erythromycin polyketide synthase dehydratase. J. Mol. Biol. 384, 941-953.
f

57) Keatinge-Clay, A.T. (2012). The structures of type I polyketide synthases. Nat. Prod. Rep. 29, 1050-1073.

58) Keatinge-Clay, A.T. (2016). Stereocontrol within polyketide synthase assembly lines. Nat. Prod. Rep. 33, 141-149.

59) Mattheus, W. Gao, L.J. Herdewijn, P. Landuyt, B., Verhaegen, J. Masschelein, J. Volckaert, G. and Lavigne, R. (2010). Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. Chem. Biol. 17, 149-159.

60) Moebius, N. Ross, C. Scherlach, K. Rohm, B. Roth, M. and Hertweck, C. (2012). Biosynthesis of the respiratory toxin bongkrekic acid in the pathogenic bacterium Burkholderia gladioli. Chem. Biol. 19, 1164-1174.

61) Moldenhauer, J. Chen, X. Borriss, R. and Piel, J. (2007). Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. Angew. Chem. Int. Ed. 46, 8195-8197.

62) Mugnai, M.L. Shi, Y. Keatinge-Clay, A.T. and Elber, R. (2015) Molecular dynamics studies of modular polyketide synthase ketoreductase stereospecificity. Biochemistry 54, 2346-2359.

63) Murshudov, G.N. Skubak, P. Lebedev, A.A. Pannu, S.N. Steiner, R.A. Nicholls, R.A. Winn, M.D. Long, F. and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355-367.

64) Otwinowski, Z. and Minor, W. (1997). Processing X-ray diffraction data

collected in oscillation mode. Methods in Enzymology. Vol 276: Macromolecular Crystallography part A. 307-326. Carter, C.W. Jr., Sweet, R.M., Eds. Academic Press (New York).

65) Perlova, O. Gerth, K. Kaiser, O. Hans, A. and Müller, R. (2006). Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain Sorangium cellulosum So ce56. J. Bacteriol. 121, 174-191.

66) Piasecki, S.K. Zheng, J. Axelrod, A.J. Detelich, M.E. and Keatinge-Clay, A.T. (2014) Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective α- and β-ketoreduction. Proteins 9, 2067-2077.

67) Pöplau, P. Frank, S. Morinaka, B.I. and Piel, J. (2013) An enzymatic domain for the formation of cyclic ethers in complex polyketides. Angew. Chem. Int. Ed. 50, 13215-13218.

68) Reid, R. Piagentini, M. Rodriguez, E. Ashley, G. Viswanathan, N. Carney, J. Santi, D.V. Hutchinson, C.R. and McDaniel, R. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry 42, 72-79.

69) Sandmann, A. Sasse, F. and Müller, R. (2004). Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. Chem. Biol. 11, 1071-1079.

70) Schneider, K. Chen, X. Vater, J. Franke, P. Nicholson, G. Boriss, R. and Süssmuth, R.D. (2007). Macrolactin is the polyketide biosynthesis product of the pks2 cluster of Bacillus amyloliquefaciens FZB42. J. Nat. Prod. 70, 1417-1423.

71) Straight, P.D. Fischbach, M.A. Walsh, C.T. Rudner, D.Z. and Kolter, R. (2007) A singular enzymatic megacomplex in Bacillus subtilis. Proc. Natl. Acad. Sci. USA 104, 1506-1509.

72) Skiba, M.A., Sikkema, A.P., Fiers, W.D., Gerwick, W.H., Sherman, D.H., Aldrich, C.C., and Smith, J.L. (2016) Domain organization and active site architecture of a polyketide synthase C-methyltransferase. ACS Chem. Biol. 11, 3319-3327.

73) Teta, R. Gurgui, M. Helfrich, E.J. Künne, S. Schneider, A. Van Echten-Deckert, G. Mangoni, A. and Piel, J. (2010) Genome mining reveals trans-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum bacteroidetes. Chembiochem 11, 2506-2512.

74) Valenzano, C.R. You, Y.O. Farg, A. Keatinge-Clay, A.T. Khosla, C. and Cane, D.E. (2010) Stereospecificty of the dehydratase domain of the erythromycin polyketide synthase. J. Am. Chem. Soc. 132, 14697-14699.

75) Weber, T. Laiple, K.J. Pross, E.K. Textor, A. Grond, S. Welzel, K. Pelzer, S. Vente, A. and Wohlleben, W. (2008) Molecular analysis of the kirromycin gene cluster revealed by β-alanine as a precursor of the pyridine moiety. Chem. Biol. 15, 175-188.

76) Weissman, K.J. (2006). The structural basis for docking in polyketide synthases. ChemBioChem. 3, 485-493.

77) Whicher, J.R. Smaga, S.S. Hansen, D.A. Brown, W.C. Gerwick, W.H.

Sherman, D.H. and Smith, J.L. (2013). Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. Chem Biol. 20, 1340-1351.

78) Yan, J. Gupta, S. Sherman, D.H. and Reynolds, K.A. (2009). Functional dissection of the pikromycin polyketide synthase into monomodules by using a matched pair of heterologous docking domains. ChemBioChem 10, 1537-1543.

79) Zeng, J. Wagner, D.T. Zhang, Z., Moretto, L. Addison, J.D. and Keatinge-Clay, A.T. (2016). Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases. ACS Chem. Biol. 11, 2466-2474.

80) Zheng J. Taylor, C.A. Piasecki, S.K. and Keatinge-Clay, A.T. (2010). Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. Structure 18, 913-922.

81) Zheng, J. Fage, C.D. Demeler, B. Hoffman, D.W. and Keatinge-Clay, A.T. (2013). The missing linker: a dimerization motif located within polyketide synthase modules. ACS Chem. Biol. 8, 1263-1270.

82) Zheng, J. Piasecki, S.K. and Keatinge-Clay, A.T. (2013). Structural studies of A2-type modular polyketide synthase reveals features controlling α-substituent stereochemistry. ACS Chem. Biol. 20, 1964-1971.

83) Smith, G. Hahn, F. (2014) A dehydratase domain in ambruticin biosynthesis displays additional activity as a pyran-forming cyclase. Angew Chem Int Ed Engl. 53, 14240-4.

84) Hanna, L. Marcio V. B. D. Williams, S. R. Hong, H. Oliveira, L. G. and Leadlay, P. F. (2015) Enzymology of Pyran Ring A Formation in Salinomycin Biosynthesis. Angew Chem Int Ed Engl. 54, 13622–13625.

85) Pöplau, P. Frank, S. Morinaka, B. I. and Piel, J. Angew. (2013) An enzymatic domain for the formation of cyclic ethers in complex polyketides. Angew Chem. Int. Ed. 52, 13215–13218.

86) Ueoka, R. Uria, A. R. Reiter, S. Mori, T. Karbaum, P. Peters, E. E. Helfrich, E. J. N. Morinaka, B. I. Gugger, M. Takeyama, H. Matsunaga, S. and Piel, J. (2015) Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. Nat. Chem. Biol. 11, 705–712.

87) Akey, D. L. Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H., and Smith, J. L. (2010) Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. Structure 18, 94–105.

88) Wagner D.T. Zeng J. Bailey C.B. Gay D.C. Yuan F. Manion H.R. and Keatinge-Clay, A.T. (2017) Structural and Functional Trends in Dehydrating Bimodules from trans-Acyltransferase Polyketide Synthases. Structure 25, 1045-1055.

89) Eustáquio, A.S. Janso, J.E. Ratnayake, A.S. O'Donnell, C.J. and Koehn, F.E. (2014) Spliceostatin hemiketal biosynthesis in Burkholderia spp. is catalyzed by an iron/ α -ketoglutarate-dependent dioxygenase. Proc Natl Acad Sci U S A. 19, 3376-85.

90) Fiers, W.D. Dodge, G.J., Sherman, D.H., Smith, J.L. and Aldrich, C.C. (2016) Vinylogous Dehydration by a Polyketide Dehydratase Domain in

Curacin Biosynthesis. J Am Chem Soc 138, 16024-16036.

91) Gay, D. You, Y.O. Keatinge-Clay A.T. and Cane DE. (2013) Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. Biochemistry 52, 8916-28.

92) Matilla, M.A. Stöckmann, H. Leeper, F.J., and Salmond, G.P. (2012) Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin. J Biol Chem 287, 39125-38.

93) Kusebauch, B. Busch, B. Scherlach, K. Roth, M. and Hertweck, C. (2010) Functionally distinct modules operate two consecutive α,β-β, γ double-bond shifts in the rhizoxin polyketide assembly line. Angew Chem Int Ed Engl 49, 1460-4.

94) Keatinge-Clay, A.T. (2017) Polyketide Synthase Modules Redefined. Angew Chem Int Ed Engl 56, 4658-4660.

95) Zhang, L. Hashimoto, T. Qin, B. Hashimoto, J., Kozone, I. Kawahara, T. Okada, M. Awakawa, T. Ito, T. Asakawa, Y. Ueki, M. Takahashi, S. Osada, H. Wakimoto, T. Ikeda, H. Shin-Ya, K. Abe, I. (2017) Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. Angew Chem Int Ed Engl 56, 1740-1745.

96) Zhang, F. He, H.Y. Tang, M.C. Tang, Y.M. Zhou, Q. and Tang, G.L. (2011) Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. J Am Chem Soc 133, 2452–2462.

97) Irschik, H. Jansen, R. Gerth, K. Hofle, G. and Reichenbach, H. (1987) The corallopyronins, new inhibitors of bacterial RNA synthesis from Myxobacteria. J. Antibiotic 40, 7-13.

98) Keatinge-Clay, A.T. Crystal structure of the erythromycin polyketide synthase dehydratase. (2008) J Mol Biol 384, 941-53.

99) Xavier, R. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server Nucleic Acids Research 42, 320–324.

100) P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. Headd, J., Hung, L.-W. Kapral, G. J. Grosse-Kunstleve, R. W. McCoy, A. J. Moriarty, N. W. Oeffner, R., Read, R. J. Richardson, D. C., Richardson, J. S. Terwilliger T. C. and. Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst 66, 213-221.

101) Langer, G. Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols 3, 1171-1179.

102) M. D. Winn et al. (2011) Overview of the CCP4 suite and current developments. Acta. Cryst. 67, 235-242.

103) Murshudov, G.N. Vagin, A.A. and Dodson, E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood method. Acta Cryst 53, 240-255.